A UNIFIED PROBABILISTIC MODEL FOR PREDICTING OCCUPANCY, DOMESTIC HOT WATER USE AND ELECTRICITY USE IN RESIDENTIAL BUILDINGS

Jean Rouleau[a], Alfonso Ramallo-Gonzalez[b], Louis Gosselin[a*], Pierre Blanchet[c], Sukumar Natarajan[d]

[a]Department of Mechanical Engineering, Université Laval, Quebec City, QC, Canada

[b]Department of Computer Science, Universidad de Murcia, Murcia, Spain

[c]Department of Wood and Forest Sciences, Université Laval, Quebec City, QC, Canada

[d]Department of Architecture and Civil Engineering, University of Bath, Bath, UK

Abstract

A novel strategy that combines separate probabilistic models developed by other researchers into a unified model for generating schedules of active occupancy, domestic hot water (DHW) use, and non-HVAC electricity use in multiple residences with a 10-minute resolution for every day of the year is described. A variety of new model functions are introduced in order to generate stochastic predictions for each of numerous residences at once, to enforce appropriate variability of behaviors between dwellings and to ensure that domestic hot water and electricity use are coincident with occupancy. The separate models used in this paper were previously developed for the US and the UK; in the unified model, scaling factors were added to these models to adjust the predictions so as to better agree with national aggregated data for Canada. The unified model was validated with measurements of domestic hot water use and electricity consumption from the 40 residential units of a social housing building in Quebec City, Canada. The behavior of occupants in the case study building was simulated 100 times in order to validate the outputs of the unified model. Goodness-of-fit tests applied to each of these simulations showed that the fit between simulated and measured dwelling-per-dwelling distributions was acceptable for 97% of the DHW consumption profiles and for 92% of the electricity consumption profiles. However, there remain discrepancies between simulations and measurements, such as an overestimation of the DHW and electricity consumption in the morning.

* Corresponding author: Louis.Gosselin@gmc.ulaval.ca; Tel.:+1-418-656-7829; Fax: +1-418-656-5343.

Keywords: Occupant behaviour; Domestic electricity consumption; Domestic hot water use; Energy modelling; Stochastic model; Social housing

1. Introduction

Up to 40% of the global energy demand comes from buildings [1], in part as a result of inefficient design, construction and operational practices. Although low energy design and construction approaches have achieved some success, it is known that poor operational practices could compromise design performance targets by a factor of at least two [2]. Reduction of energy consumption therefore needs to come not only from using improved design and construction technologies, but also from recognizing the impact of occupant behavior [3][4]. Yet, despite detailed investigations of occupant behavior and its impact on energy demand [5]–[7], variations in occupant behavior are scarcely considered in practice in building modeling.

Many user behaviors affect building energy performance, including: the number of active occupants present, the use of electrical appliances, the use of domestic hot water (DHW) appliances, the use of lighting, the control of the heating cooling and ventilation systems, the control of window openings, and the control of blinds. At present, the industry normally uses static schedules (i.e. typical daily schedules that are repeated over the years) to represent these actions in energy simulations. Although these deterministic schedules still have their place in building simulation depending on the application (the final report of the EBC Annex 66 of the IEA has a chapter on the importance of fit-for-purpose in occupant behavior simulation [8]), they have their shortcomings. With such an approach the amount of heat, DHW and electricity used in a specific building at a given time is fixed and corresponds to an "average" expected behavior [8][9]. In reality, different individuals have different preferences and hence adopt different behaviors. Consequently, any particular building may have a range of possible energy consumption levels instead of the single value obtained with static schedules. Hence, it is not surprising that great differences are often observed between the predicted energy consumption of a building and its actual energy consumption. This so-called "energy gap" is most frequently due to occupant behavior [11].

Another way to depict occupant behavior in building simulations that has the potential of fixing this issue is the use of probabilistic models [12]–[15]. Since these models are based on probabilities instead of a purely deterministic approach, they allow the representation of more diverse occupant behaviors. These stochastic models allow new ways of performing building designs. For example, Ramallo-González et al. initiated the concept of robust optimization of low-energy buildings [16]. These variations lead to different levels and patterns of energy end uses and thus can capture the wide range of possible annual energy consumption of a building.

Most existing probabilistic occupant behavior models were built upon country dependent data [17]–[19]. Since occupant behavior depends on socio-economic and psychological factors, cultural differences can lead to different occupant behaviors, implying that occupants in different countries might act differently [12][19][20]. Consequently, most existing probabilistic occupant behavior models cannot be employed straightforwardly all around the world. One solution to this problem would be to replicate in each country the extensive monitoring process required for the development of these models in order to obtain country-specific calibrated models. Sometimes, the required data is readily available in databases [22], but this is not the case for most countries such as Canada. Despite the evident reliability and precision provided by extensive field surveys, it should be recognized that this approach is also quite cumbersome since surveys are very time-consuming and expensive to perform. Nevertheless, even when precise occupant behavior pattern is unknown, probabilistic models can offer the advantage of generating a range of plausible profiles to be considered for design or other purposes.

Another important limitation is that most probabilistic occupant behavior models found in the literature have been developed independently, and focus on individual issues (i.e.: either occupancy, or DHW use, or window openings). Consequently, a building professional may use one model to predict the occupancy in a building, then use a different model to predict the use of DHW. It would be substantially easier for users to employ one unified model instead of relying on multiple unique models with varying methodologies and differing nomenclatures.

This study investigates the potential and limits of a unified model that predicts the number of occupants, domestic hot water (DHW) use, and non-HVAC electricity use in multiple residences with a 10-minute resolution for every day of the year. Section 2 details the methodology employed to build the "integrated occupant behavior". Recognized probabilistic occupant behavior models were merged together. These original models were developed independently using data from the US, Canada and the UK. In addition to developing a unified method, this study introduces scaling factors to adjust the predictions so as to better agree with national aggregated data for Canada, and measured data from a social housing building in Quebec City. The model developed in this study could be used for other scenarios, but would need to use appropriate inputs. The model was implemented in MATLAB [23] and was primarily developed with the idea of representing occupant behavior in energy simulations of multi-residential buildings at the predesign or design stage, which dictated the required level of details and accuracy. The model could also be used for other applications, such as for predictive control or demand-side management. For example, a methodology to size the DHW system in an apartment building was developed based on an occupant behavior model in [24]. Section 3 discusses the limits of the approach that was used and the validity and precision of the model, by comparing its outputs with measurements obtained from a multi-residential building in Quebec City, Canada. Both aggregated and disaggregated demands were analyzed (Sections 3.1 and 3.2). The effects of the modifications brought to the existing occupant behavior models on the accuracy of the unified model were then thoroughly studied (Section 3.3).

2. Occupant behavior model

This section presents the methodology used to develop the unified probabilistic occupant behavior model that is described and validated in this paper. The model predicts three behaviors: the number of active occupants in each of multiple residences, the DHW consumption in each residence, and the non-HVAC electricity consumption in each residence. The model extends from work documented in a previous conference paper [25]. Each of the predicted behaviors interacts with each other to ensure that the generated outputs are consistent. The flowchart in Fig. 1 exhibits the relationships between these behaviors. The number of dwellings and the number of days must first be specified. Other important parameters that can drive variability of energy consumption such as energy price, socioeconomical status and appliance ownership are already considered by the model with the use of probability functions that compute the type of occupants in each simulated dwelling, so the users of the unified occupant behavior model do not need to provide such information. The origin of these probability functions are discussed later in the paper. By adapting these inputs, the model could also be useful for other scenarios not tested in this study. The blue boxes in Fig. 1 represent the internal parameters within the model that have to be changed so to adapt the model for a specific country.

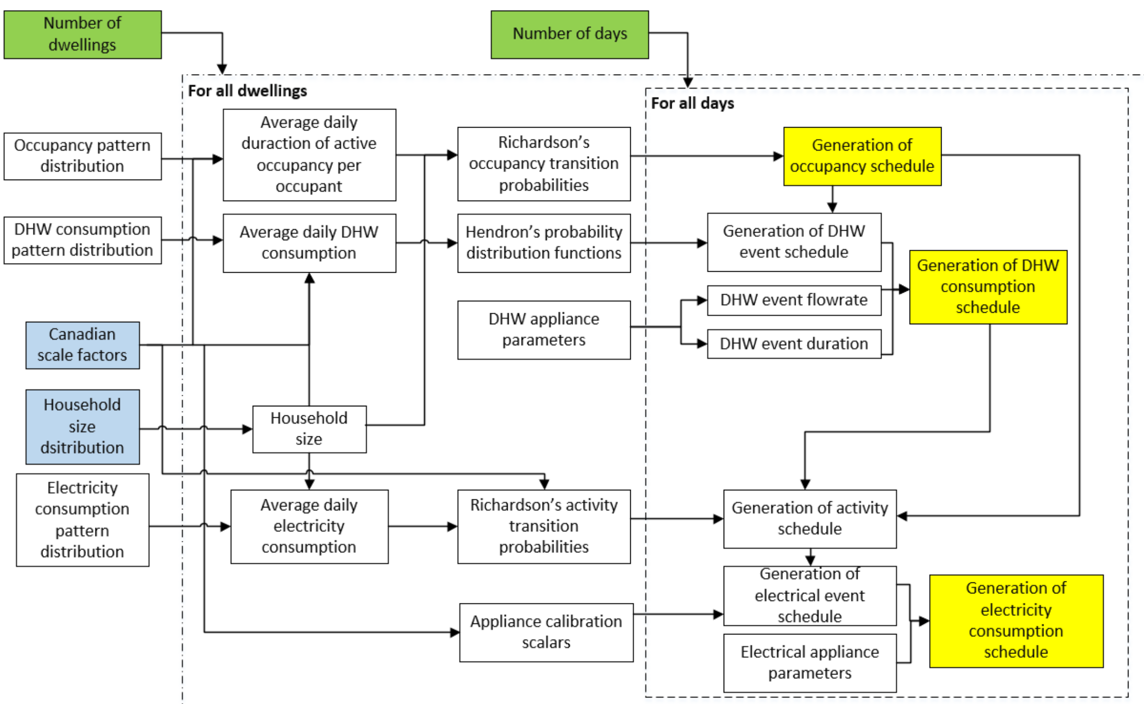

Figure 1: Architecture of the occupant behavior model showing the relationship between all components. Green boxes refer to inputs that have to be provided by the model user. Blue boxes are the building/country specific data whereas yellow boxes are the outputs of the model.

2.1 Active occupancy model

The initial step of the model is to find when occupants are active in their home. For its simplicity, the stochastic daily occupancy profiles generator developed by Richardson *et al.* [17] was chosen to serve as the basis for the active occupancy model. Active occupants are defined here as occupants that are physically present and not sleeping. Richardson's model employs a first-order Markov-chain Monte Carlo method [26]. The number of active occupants at a given time step depends only on the number of active occupants at the preceding time step, the day of the week, and the hour of the day. Richardson's model uses a 10-minute resolution, meaning that there are 144 time steps in a day. The probability of changing from one state (i.e., number of active occupants) to another is different for each of these time steps. These probabilities are logged in "transition probability matrices" that are based on a survey of 20,000 weekly UK household journals [27].

Three additions to Richardson's model were incorporated. First, the possibility of allowing the model to choose the household size of each simulated dwelling was included. In Richardson's model, the user must provide the household size. In the present model, the household size can be generated randomly based on a probability distribution of the given country (in our case from Canadian household statistics [28]). Note that this step is not mandatory if one already knows the household size of the dwellings.

The second adjustment modifies Richardson's model to fill in unknown parameters for one country using data from a different country. Researchers have developed occupancy models that are similar to Richardson's in the US [29], Spain [30] and Sweden [31]. The center for Time Use Research in Oxford have uploaded data files that contain time use information from dozens of countries [22]. However, for some of these countries (Canada being one of them), the available time use information provides the number of minutes spent by citizens on various activities, but not the starting time of these activities. It is thus impossible with that data to find precisely at what time occupants were actively at home, preventing the replication of Richardson's methodology to create occupancy simulator for those countries. However, it is possible to compute the aggregated daily amount of time during which a person is actively at home. Knowing this data for two countries, it is possible to calculate a scale factor to adapt an occupancy model developed in one country so as to represent occupancy in a different country. Referring to the case of the UK, time-use survey overviews say that British citizens spend on average 1,003 minutes per day in their home and sleep for 476 minutes, meaning that they are active in their dwellings for 527 minutes per day [27]. In Canada, these numbers are 990 minutes at home and 498 minutes of sleep; consequently for this study 492 minutes of active occupancy was used. [32]. Therefore, Canadians spent on aggregate 35 fewer minutes per day awake at home than British – an average reduction of 6.6% of active occupancy. For this scaling approach to be appropriate, one has to assume that the lifestyle in the two countries considered is not too dissimilar.

Any time a random number is drawn to find the number of active occupants for the next time step, the number is multiplied by a scale factor that ensures that occupancy respects national aggregated data. The model was run 1,000 times after the application of this scale factor for a household during a weekday and a weekend day. This number of simulations was chosen based on the work of McKenna *et al*., who showed with a similar model that negligible variations of aggregated results are found after 1,000 simulations [33]. It showed that active occupancy lasts for 473.0 minutes during weekdays, 539.2 minutes during weekend and thus as expected 492.0 minutes per day on average. The main effect that this change had on the aggregated occupancy daily schedules was to reduce slightly the probability of occupants being active throughout the day. Therefore, this scaling methodology relies on the assumption that apart from the total time of active occupancy, people from the two countries that are compared are likely to follow similar occupancy patterns (i.e. waking up at the same time of the day and likewise for going to work, coming back home and going to sleep). It is clear that the assumption that the occupancy pattern in a country can serve as the basis for developing the occupancy pattern in another country might not be true if the two countries are too dissimilar. Evidently, when one would already have access to TUS data or to a specific occupancy model for the country of interest, it would be preferable to refer to this data. However, when such detailed information in unavailable, the proposed methodology could be considered, and in that case, the scaling is a simple and convenient way to adapt the occupancy profiles with the available information.

The final modification accounts for diversity in occupancy patterns between different households. Families have different needs and live through different situations, meaning that some households have individuals at home more often than other households. To reproduce this "dwelling-to-dwelling" variability, the model employs a probability distribution to assign an average daily occupancy duration to each dwelling. This methodology does not necessarily cover all possible occupancy patterns, but it captures a more realistic diversity of occupied hours per dwelling. The chosen probability distribution assumes that the average amount of time spent at home for a dwelling follows a normal distribution since no indication were found as to what distribution law should be used. The mean of the distribution is set to one so that its introduction in the model will not affect the aggregated occupancy. The standard deviation was computed with results from Aerts *et al.*, who found that people who are mostly absent from home spend approximately 240 minutes per day at home while those mostly at home stay there 720 minutes when they clustered households in seven distinct groups according to their occupancy profiles [34]. This work was made in Belgium, where the average active occupancy is 493 minutes per day [34]. The standard deviation of 114 minutes was chosen for the normal distribution of occupied daily hours per dwelling so that the range of values agrees with Aerts' data. This standard deviation is equal to 23% of the mean value. Therefore, for every household, the scale factor in the model is multiplied by a random parameter which follows a normal distribution with a mean value of $\mu = 1$ and a standard deviation of $\sigma = 0.23$.

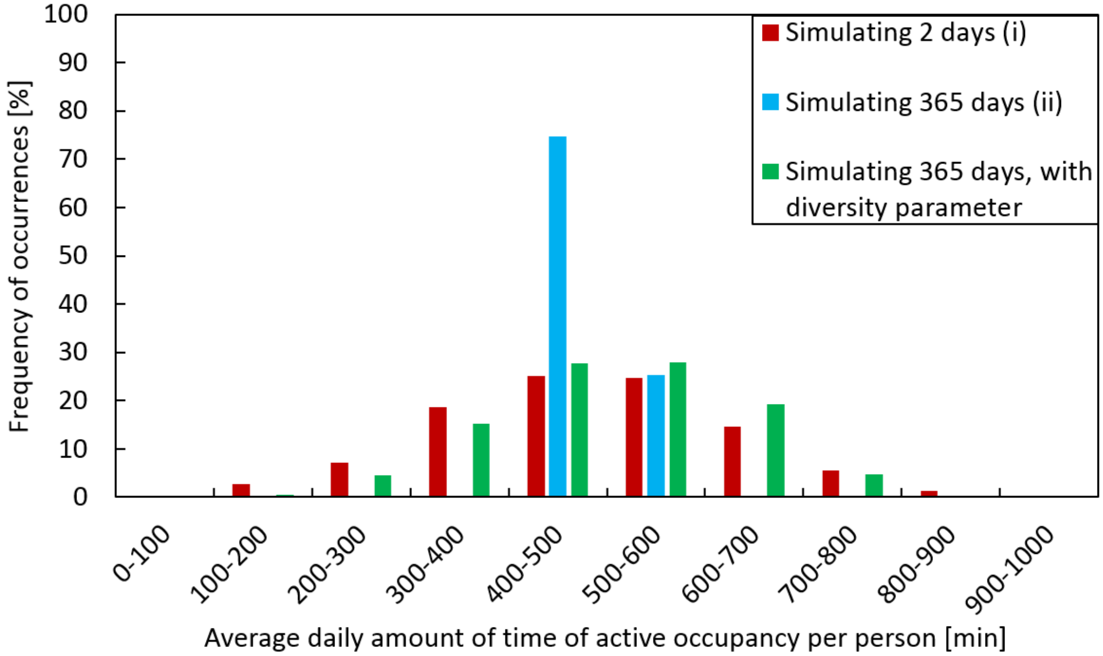

Figure 2: Distribution of the average daily amount of time of active occupancy per person in 1,000 simulated dwellings according to different models. The average daily amount of time of active occupancy should be between 240 and 720 minutes.

The methodology was used to obtain annual profiles for 1,000 dwellings. Their total time of active occupancy was then separated into distinct bins of 100 minutes per day per person. Fig. 2 shows the resulting distribution and compares it to the one obtained by repeating this process for two other simulation strategies that do not employ a distribution to infer "dwelling-to-dwelling" diversity: (i) simulating one weekday and one weekend day for a dwelling and replicating them over a year (use the obtained weekday schedule for 261 days and the weekend one for the remained 104 days) and (ii) simulating 365 days (use a different simulated schedule day after day) without inducing diversity between the households. Strategy (i) should yield "dwelling-to-dwelling" variability in occupancy patterns, but practically no "day-to-day" variability since the same days are repeated over and over again. For strategy (ii), it is the opposite – occupancy schedules are different day after day, but all households should have similar aggregated occupancy behaviors since no diversity was enforced. This is shown in Fig. 2 where the latter option leads to a very narrow distribution that is not close to the target "dwelling-to-dwelling" diversity (240 to 720 minutes of active occupancy) found from Aerts' study. The "simulating two days" solution tends on the other hand to overrate the diversity of occupancy as a non-negligible proportion (10.7%) of the dwellings are outside the target “dwelling-to-dwelling” diversity. This option yields a standard deviation of 148 minutes per day, which overestimates the target of 114 minutes by 29.8%. The average value obtained from multiple draws is quite variable for small numbers of draws but will converge towards a specific value for a large number of draws. This explained why the “simulating 365 days” strategy greatly underestimates the diversity of occupied hours whereas the “repeating 2 days over a year” strategy overcompensate. These results are based on the assumption that the probability distribution used in the model to enforce diversity in occupancy patterns is accurate.

2.2 Domestic Hot Water (DHW) model

Few probabilistic DHW models that generate volumetric consumption are available in the literature [17][34][35]. Most of the DHW models are integrated in thermal domestic demand models that compute the thermal demand for DHW. These models use a range of methods such as non-homogeneous Markov chains [32][36][37], time-series [39], probability density functions [15] or neural network [40] to predict the heat demand due to the consumption of water.

A popular and easy-to-use model is the yearly DHW event schedule generator developed by Hendron *et al.* [18], [41]. This model generates an annual volumetric DHW profile for a single dwelling by dividing DHW consumption into five types of water appliances (shower, bath, sink, clothes washer and dishwasher). Each appliance has a daily probability density function (PDF) that determines the probability that the appliance is involved in a hot water event at each hour. These PDFs were computed with datasets coming from two monitoring studies in the United States [41]. When the model predicts a hot water event, the volumetric consumption is calculated by multiplying the duration of the event with the flowrate at which water is consumed. These two variables are randomly chosen according to different PDFs that are specific to the five hot water appliances. This model is based on data coming from one country and, like Richardson’s occupancy model, might not adequately represent the DHW demand patterns in other countries.

Six modifications were implemented to adapt Hendron's model for the model described in this paper. First, a linear interpolation was made to adjust the hourly resolution of the start-time PDFs from hourly resolution to the 10-minute resolution used in the model. Second, a calibration scalar was added to account for the household size. There should be more hot water events in dwellings that have large household size and vice versa. As suggested by other studies [42]–[44], a linear scaling with a slope of 35 litres per person, divided within the five appliances, is used for this calibration. This slope is equal to the value used by the Canadian building simulation software HOT2000 [45].

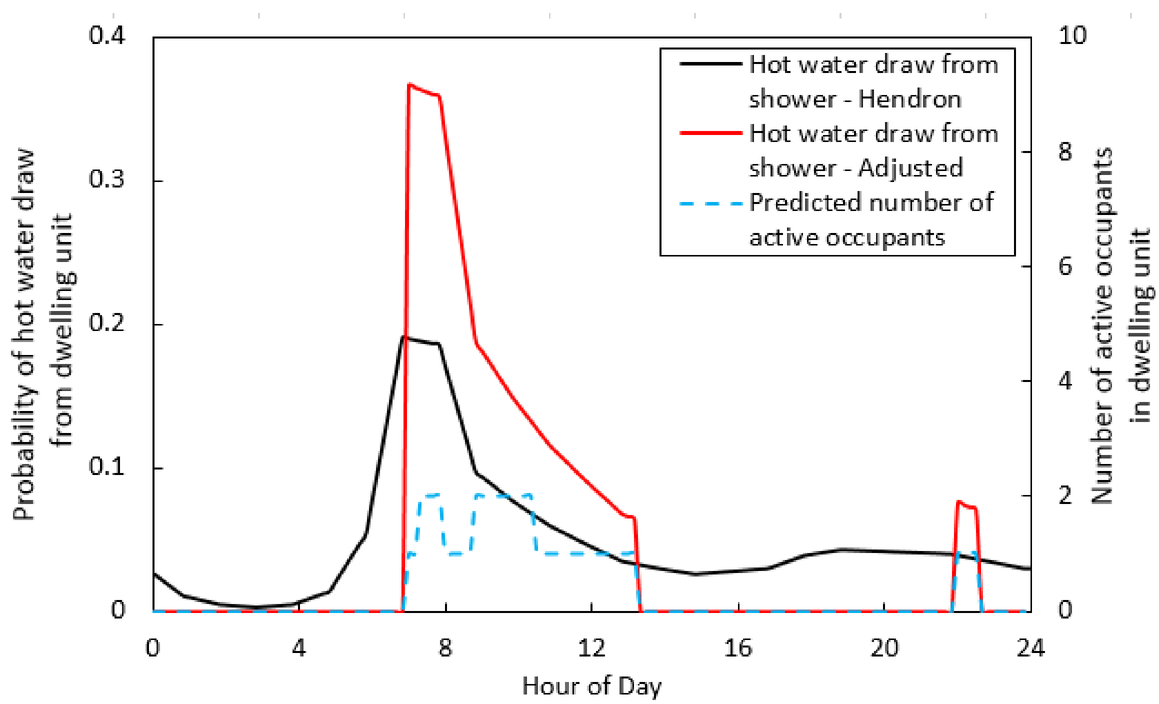

Figure 3: Modification made to the probability density function of a shower event to account for active occupancy.

The third modification links DHW consumption to occupancy. The shower, bath and sinks cannot use DHW when there are no occupants active in the building. In addition, there should be more DHW consumption when there are many active occupants in the dwelling. Therefore, for all time steps, the PDFs are multiplied by the projected number of active occupants to increase the probability curves in time steps with high occupancy. The area under the curve of the new PDFs must be equal to the initial ones to ensure that the daily total DHW use is unaffected by this change. The modified functions are thus multiplied by a correction factor that is equal to the ratio between these two areas. Fig. 3 offers a graphical example of this procedure for the probability of using the shower during a single day. The aggregation achieved by simulating 1,000 different days is shown in Fig. 4. If active occupancy (blue curve) had no influence, the probability curve before the fitting with occupancy (black curve) would perfectly be superimposed with the aggregated function generated after the fitting (red curve). The morning peak in the aggregated PDF happens an hour later than in the previous function, probably due to the British origins of the occupancy model versus Hendron's model which was developed for the USA. In the evening, since it is the peak period for active occupancy, there is an increase in the probability of a shower event. The integration of the black and red curves provides identical values, demonstrating that this treatment is only affecting the timing of events and not the overall quantity of events.

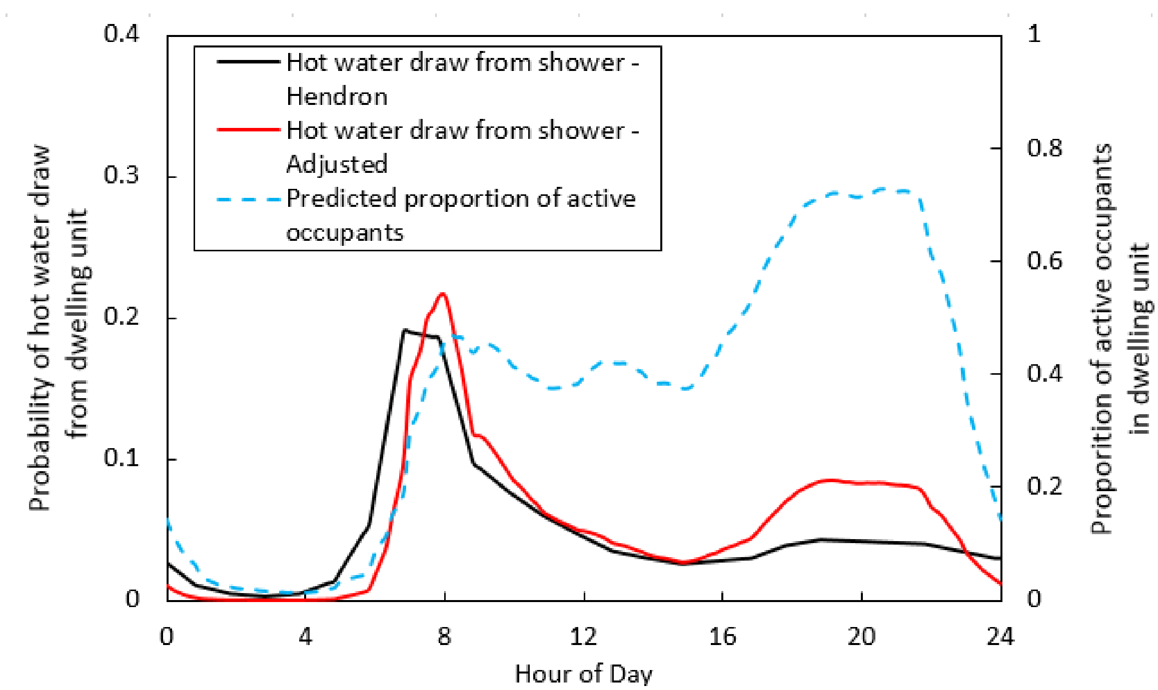

Figure 4: Aggregated start-time probability density function for the shower before and after accounting for active occupancy.

The fourth adjustment scales Hendron's model from American to Canadian data (see Table A2). A scale factor reduces the PDFs that are used for the duration of hot water events since Americans and Canadians have slightly different DHW consumption levels. The fifth modification is another scale factor that decreases the flowrate to account for low-flow devices (showerheads, dishwashers, washing machines and sinks) that are getting more widespread. A reduction factor of 20% was selected based on an analysis of retrofits in [46]. This factor is applied to all appliances except for the bath.

The sixth and final change to Hendron's model was the consideration of diversity in the level of consumption between dwellings. To do so, a scalar is drawn from a "diversity" PDF that is based on a monitoring study [42]. This study provides the distribution of daily DHW consumption of 119 households, ranging from an average of 12.5 L/day to 612.5 L/day with a mean value of 172.0 L/day. Part of that variability is due to the number of occupants forming these households, but the study also gives the distribution of occupancy in the monitored dwellings in addition of a best fit equation to find the average daily DHW consumption in L/day from the household size:

$$V_{DHW} = 39 \times \#Occ + 17 \tag{1}$$

where #Occ is the number of occupants living in the dwelling. By combining this best fit equation with the occupancy distribution, it is possible to find what the distribution of DHW consumption would be if every occupant asked for the same volume of water. Fig. 5 compares this "household size based" distribution with the one actually measured in the 119 homes. It is clear that the measured distribution is larger than the one predicted strictly with the household sizes – more dwellings have an average consumption below 100 L/day and above 300 L/day. This is suspected since people have different habits and some use more DHW than others. A random parameter has to be applied to Eq. (1) to simulate this aspect. Different distributions were tested and it was found that the log-normal distribution with a mean of $\mu = 0$ and a standard deviation of $\sigma = 0.35$ provided the best fit between the generated DHW consumption distribution and the one measured in the study. The average output of a log-normal distribution with $\mu = 0$ is 1 so this introduced parameter does not change the predicted aggregated volume of water. Therefore, in the model, each dwelling received a 'diversity' parameter from this distribution which is multiplied by the duration of hot water events to calibrate the total volume consumed by the household. This modification changes the average volume of water used per event, but not the number of events itself, i.e. heavy DHW users are considered in the model as people taking long showers, not as people taking many showers. The frequency of hot water events is already linked with the number of occupants living in the dwelling.

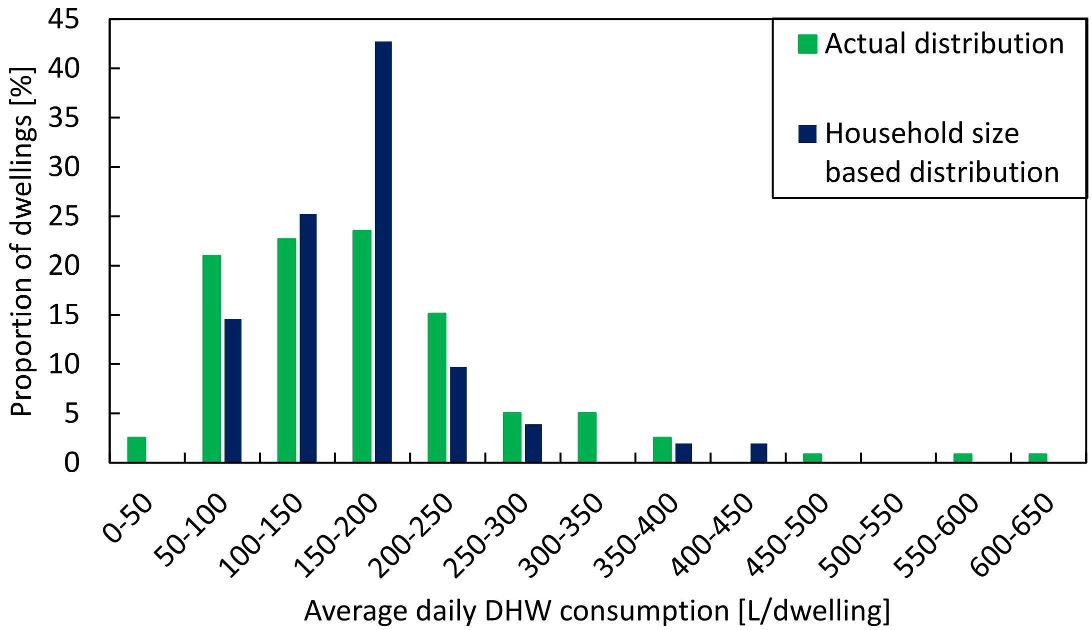

Figure 5: Comparison of the measured density of average daily DHW consumption with the one generated by only considering household sizes.

2.3 Electricity model

Several residential electricity consumption models have been created by previous researchers to predict the intensity and timing of demand and peaks and various methodologies have been proposed. For instance, Chitnis and Hunt developed a model that uses financial aspects (price of electricity, household income, appliance ownership…) as independent variables to help predict residential electricity consumption [47]. Harris and Liu included weather data (temperature, precipitation…) in their electricity consumption [48]. The type of occupants (age, gender, education) is considered in the model created by Fischer et al. [49]. “Economic” models often run into the problem of combining aggregated economic data with disaggregated load profile data, hence the recent gain in popularity of “non-economic” models that prefer to use time-use surveys as their basis [36][47]–[49]. Two of these time-use surveys based models are the ones developed by Richardson [19] and Armstrong [53], which were both taken in this paper to simulate the electricity consumption. Since it is already connected with the active occupancy model, Richardson's was taken to generate schedules for the use of electric appliances, as these schedules greatly depends on active occupancy. As for Armstrong's model, it was employed for the usage of the lighting systems. Armstrong's model has the advantage (in the context of this paper) of being based on Canadian lifestyle.

Like his occupancy model, Richardson's electricity use model relies on the Markov-Chain technique. This technique is an efficient way to model the use of electrical appliances as these appliances have two possible states (on/off). Consequently, their popularity in electricity forecasting models is not surprising [11], [27], [51]. In time-use based electricity models, Markov chains create daily schedules of activities in a building by identifying the times at which occupants switch from one activity (e.g.: cooking, laundry, watching TV) to another. The probability density functions for transition between different activities were computed from time-use survey data, as in his active occupancy model. Every individual appliance is linked to an activity so that its likelihood of being used increases once the corresponding activity is ongoing in the generated activity schedule. Contrary to Hendron's DHW model, when an appliance is seen as being activated, it is used for a constant duration with a specific power consumption since no data could be found on the variability of the duration of use of the electrical appliances considered in the model. Future iterations of the model could include this detail.

Once again, Richardson developed a model that is based on measured household electricity use in the UK and aggregated electricity use data from Canada was used to scale Richardson's model to fit with Canadian lifestyle so the predictions of the model could be validated with the data available for this specific work. Table A1 lists the aggregated amount of time that a Canadian spends on cooking, on watching TV and on household work [32]. Differences are observable between this data and the ones found in time-use surveys made in the UK [27]. The activity probabilities were multiplied by a scale factor to ensure that the aggregated results are identical to the left column of Table A1.

Table A3 contains the list of appliances that are considered in the model shown in this paper. Out of the 33 electrical appliances that are considered in Richardson's model, some were taken off. *Chest freezer*, *Fridge freezer* and *Upright freezer* were merged in one single appliance named *Freezer*. Likewise, *Tumble dryer* and *Washer dryer* became *Dryer*. *Answer machine, Cassette Player, Clock, VCR/DVD player, Cordless telephone, Fax* and *Printer* were eliminated as they either are devices that are rarely seen in dwellings today or that consume a negligible amount of energy. *Small cooking (group)* was divided in multiple end-uses: *Toaster, Exhaust fan* and *Coffee Maker*. Moreover, all appliances related to electric domestic water or space heating were not considered since this model is about the non-HVAC electricity consumption of residential buildings. Two additional devices were introduced: *Laptop computer* and *Hair dryer.*

The activity *None* in Table A3 means that the appliances do not require active occupancy to be operating. For devices that are associated with *Occupant,* there has to be at least one active occupant in the dwelling for them to be turned on. The C*lothes washer* and *Dishwasher* appliances are simulated differently since they are linked to *Domestic Hot Water*. The DHW part of the model directly identifies time steps in which these appliances are used, so there is no need for calibration scalars. The rest of the activities are the ones considered by Richardson and are simulated with the activity probabilities matrix: *Watching TV, Cooking, Laundry, Washing/Dressing, Iron and House cleaning*. The probabilities of use provided in Table A3 describe the likelihood that an appliance is operating once its corresponding activity is enabled in the activity schedule. For example, when the *Cooking* activity is happening, there is a probability of 17.2% that the hot plate is used by the occupants. For their calculations, the total number of hours of operation per year has to be computed:

$$D_i = \frac{1000E_i - 8760P_{off,i}}{P_{on,i} - P_{off,i}} \quad \text{for } i = 1, 2..., m \tag{2}$$

where E_i is the aggregated energy consumption in kWh measured in Canadian homes found in Table A3 for appliance i, $P_{on,i}$, its power consumption when operating and $P_{off,i}$, the standby consumption. Inserting proper numerical values in Eq. (2) gives, for example, a use of 168.3 hours per year for the hot plate. Knowing this duration, it is possible to find the annual number of events:

$$M_i = \frac{60D_i}{\lambda_i} \text{ for } i = 1, 2..., m \tag{3}$$

where λ_i is the event length in minutes. Continuing with the example of the hot plate, which was attributed an event length of 16 minutes, the model must produce an average of 631 events per year. To obtain the probability that people use the hot plate when cooking, the total number of time steps in which the *Cooking* activity is activated is needed:

$$N_j = \frac{365\delta_j \times 2.4}{\Delta t} \text{ for } j = 1, 2..., n \tag{4}$$

Here, δ_j represents the daily aggregated amount of time spent on activity j and Δt the model time step. δ_j is multiplied by 2.4 because according to the household size distribution, the mean household size is 2.4 occupants per dwelling. For the *Cooking* activity, Canadians cook 42 minutes per day, meaning that in the average dwelling, there is cooking for 100.8 minutes per day (36,792 minutes per year). With a time step of 10 minutes, this translates for the model into 3,679.2 time steps in which *Cooking* should be enabled. The probability that the hot plate is operating when cooking is merely the ratio between the targeted amount of hot plate events and the number of *Cooking* time steps:

$$P_i = \begin{cases} \dfrac{M_i}{N_j} & \text{if } j = \text{on} \\ 0 & \text{if } j = \text{off} \end{cases} \tag{5}$$

Hence, a probability of use of 631 / 3679.2 = 17.2% for the hot plate. The same procedure was repeated for all appliances to get the parameters displayed in Table A3.

As previously mentioned, Armstrong's electricity model, which is based on probability density functions, was used to simulate the consumption of the lighting systems. Each season has its own daily probability curve to calculate the odds of a lighting event happening. Use of lighting greatly depends on multiple building aspects, such as its localization and orientation, its window-to-wall ratio or the shading of the surrounding buildings. For the sake of simplicity, these aspects are not considered in these PDFs. The variability of lighting appliance use introduced by these aspects is assumed by Armstrong to be included in the probabilistic aspect of the model. When a lighting event occurs, the power consumption varies between 60 and 410 W and the duration of the event is selected between 5 and 120 minutes. These two parameters are selected based on a uniform random distribution. The modification made to Armstrong's model was to adapt the PDFs so they fit with occupancy profiles. The treatment applied to Hendron's model to account for occupancy was repeated for the probability curves of lighting events.

For each dwelling, a scale factor was applied to the 'probability of use' parameters for electrical appliances. This factor was defined as the product between three sub-factors: one that is due to household size $s_{\#Occ}$, another for the type of consumer $s_{consumer}$ and a final one to consider the type of building $s_{building}$:

$$s_{dwelling} = s_{\#Occ} \times s_{consumer} \times s_{building} \tag{6}$$

The 'number of occupants' sub-factor ($S_{\#Occ}$) was estimated with data taken from Statistics Canada suggesting that the relation between electricity consumption and household size has a slope of approximately 3.75 kWh/day per occupant [55]. As for the 'type of consumer' sub-factor ($S_{consumer}$), according to Armstrong the mean daily electricity use for detached houses in Canada ranges from 13.2 to 35.6 kWh/day. Unfortunately, since studies on the diversity of electricity consumption between different people are rare, it was not possible to isolate the variations of consumption that are due to the household size. Applying the methodology used to determine diversity in active occupancy, the range delimited by 13.2 and 35.6 kWh/day corresponds to a normal law with a mean value of 24.5 kWh/day and a standard deviation of 5.6 kWh/day. The standard deviation is equal to 22.9% of the mean value, and therefore for each dwelling a normal distribution with $\mu = 1$ and $\sigma = 0.229$ drives the value of the 'type of consumer' sub-factor. Once again, the distribution's unitary mean value ensures that this sub-factor does not affect aggregated results. A minimum of zero is set for this parameter so there cannot be negative consumption. Since this prescribed minimum is more than three standard deviations away from the mean, the distribution is not visibly truncated and the effect of this constraint on the mean output is negligible. The ‘type of building’ parameter is there to adapt the energy demand for apartments. All data related to electricity used so far were representative of consumption in detached single houses. Since the electricity consumption is quite larger in detached houses than in apartments (mostly due to a larger floor area and a larger set of electrical appliances), an adjustment is necessary to simulate consumption in apartments. In [56], which presents the overall energy consumption of 8,230,596 detached houses and 2,059,428 apartments in Canada, the average non-HVAC electricity consumption of an apartment is approximately 57% of the one of a detached house. If one wants to simulate detached house, the ‘type of building’ sub-factor should be set to 1, but it needs to be 0.57 for apartment units.

3. Comparison of the model with in situ measurements

The model was compared with measurements taken in a recently constructed multi-residential social housing building in Quebec City, Canada. Data measured in this building include DHW volumetric demand for each of the 40 dwellings along with the electricity consumption of eight apartments. These quantities were measured every 10 minutes. In addition to the real-time measurement of electricity for some of the dwellings, the electricity consumption of the remaining 32 dwellings was recorded every month by electricity meters. Since heat needed for space heating and DHW is provided to the building by radiators using hot water from a district heating system, the electricity consumption was used for non-HVAC purposes. Electricity used by the fans of the ventilation system were measured at the building level, but not at the dwelling level so it was not included in the electricity consumption of an apartment. The monitoring duration considered for the validation is a full year (from January 1st 2016 to January 1st 2017). This dataset was independent from the model – it was not used in the making of the model and therefore can be used for independent validation. In practice the occupant behavior model could be used before the construction of the building (e.g., for energy simulations or sizing equipment) and therefore, it would not be possible to adjust the model to fit in situ measurements.

The total population of the building during the monitoring period was 90 people (an average of 2.25 occupants per household). According to the household size distribution used in the model, this number was lower than average, but not abnormally low (22nd percentile of possible building population). For both DHW and electricity consumption, the objective of the work presented here was to achieve a model that accurately depicts stochasticity in occupant behavior while still offering satisfying aggregated results. Therefore, the validation of the model is divided in two parts. The first part checked the aggregated patterns, where the whole building consumption ass compared to aggregated results from the model. The other part of the validation will study diversity in consumption between individual households. Because no data were taken for active occupancy in the real building, this part of the model could not be directly validated. However, due to its link with the other two simulated behaviors, adequate consumption representation indirectly revealed whether the occupancy is appropriately simulated. Furthermore, it had already been shown in Fig. 2 that the active occupancy model generates satisfying results regarding aggregated national statistics.

3.1 Aggregated demand

Consecutive simulations of the same building can provide different results due to the stochastic nature of the model. To quantify the different possible levels of DHW and electricity consumption of the building, multiple simulations were performed and compared with the monitored building to obtain various overall annual profiles. The number of simulated dwellings was set to 40, the number of days to 365 and the household size distribution is identical to the one found in the real building (i.e. each simulation had a population of 90 people). The evolution of the distribution of building consumption is presented in Table 1 as a function of the number of simulations performed. The non-zero standard deviation (which refers to the deviation found from the distribution of average DHW consumption of each building simulation) demonstrates that the total DHW and electricity consumption of the building cannot be precisely known before operation due to the occupant behavior, even if the impact of every household is smoothened over 40 dwellings. After 100 simulations (translating into a total of 4,000 simulated dwellings), the average daily DHW use and electricity demand are respectively 134.8 litres per dwelling and 13.86 kWh per dwelling. A consumption level of 134.8 litres corresponds to a reduction of 40% from the value provided by National Resources Canada in 2012 (225 litres; see Table A2) for the average hot water consumption in a Canadian dwelling [57]. This significant drop between the model and the expected value can be explained by the small number of occupants in the building and by the installation of water saving devices. In another recent monitoring study in Canada, an average demand of 172 litres per day was measured over a sample of 119 homes that had a mean household size of 3.83 people [42]. Therefore, it is not aberrant that the level of consumption in the model is lower than the value reported by National Resources Canada. In fact, in the case study building, the average daily consumption of hot water during the monitoring period was 131.2 litres per apartment. In Fig. 6a, the distribution of the DHW consumption in the building obtained with the 100 simulated profiles is illustrated. Since the amount of DHW use in the validation data falls into the distribution generated by the model, it appears that the model is in agreement with the case study building for the total amount of hot water use.

Table 1: Variability of the DHW consumption and electricity use profiles as a function of the number of profiles generated

Number of profiles generated	**Domestic hot water** $\left[\frac{\mathbf{L}}{\mathbf{day \cdot dwelling}}\right]$		**Electricity** $\left[\frac{\mathbf{kWh}}{\mathbf{day \cdot dwelling}}\right]$	
	Average	**Standard deviation**	**Average**	**Standard deviation**
1	135.1	-	13.71	-
5	134.4	6.1	14.37	0.80
10	136.0	5.4	14.17	0.66
25	135.5	5.7	13.93	0.67
50	135.2	6.8	13.87	0.62
75	134.6	6.7	13.89	0.57
100	134.9	7.0	13.86	0.54

The distribution of electricity demand computed by the model is also shown (Fig. 6b). The average electricity consumption for a dwelling in the monitored building is 14.81 kWh per day. This figure shows that the measured electricity consumption falls within the values given by the model, with a tendency to be closer to high values.

Figure 7 compares the simulated mean daily DHW and electricity profiles throughout the year for all dwellings generated in the 100 simulations with the average profiles found in the validation data. The shaded area around the simulation curves provide the variations seen between all simulations – the area is bounded by the 5th and 95th percentiles observed from the 100 aggregated simulated profiles at every hour of the day. Consumption of hot water and electricity during the night is lower in the model than in the measurements, but the model overrates the morning peak from 7AM to 10AM – it is the only period of the day where the measured curve is out of the range generated by the simulations. After 10AM, the aggregated patterns provided by the model closely follow the ones of the case study building. Nonetheless, measured and simulated profiles have similar general behaviors: low-consumption in the early hours, followed by an increase in the morning to a level of consumption that is mostly constant until the evening peak happens. The only large difference between simulations and measurements is the morning DHW consumption. Simulations predict a peak with a consumption rate of nearly 12 litres per hour that is not happening in the monitored building. It can be argued that the occupants living in the case study building do not follow a "typical" daily DHW schedule as morning peaks are seen in most DHW monitoring studies [43]. For instance, in the previously mentioned monitoring study made in Canada [42], the consumption of hot water between 6AM and 10AM represents 28.3% of the total daily DHW demand whereas in the building used in this paper, this value goes down to 18.8%. In the simulated profiles produced by the model, 23.5% of the DHW consumption is made in that morning period. A possible explanation to this unusual behavior in the monitored building is that due to a high proportion of children, baths are more often taken in the evening instead of in the morning. Another reason for the differences might be that the modeling of active occupancy is not "perfect". Since the occupancy in the simulations is based on British schedules, there could be some errors in the representation of Canadian occupancy patterns. For example, the increase of consumption in the morning happening approximately one hour earlier in the validation data versus in the simulations can be due to Canadians waking up on average an hour earlier than British, but at this point no clear report in literature was found to confirm this assumption. A similar observation can be made for electricity – the simulation results predict more consumption between 7AM to 9AM than what is seen. Again, the metered profile slightly differs from what is seen in other electricity monitoring analyses, with a proportion of 6.1% of electricity being consumed between 7AM to 9AM. Two different samples of houses in Canada (one of 29 households in Nova Scotia and the other of 22 households in Ottawa) have a proportion of approximately 8.0% and 8.3% of electricity consumed during this period of the morning [58]. Larger samples in Europe have also yielded a fraction around 8% [56][57]. The model predicts on average that 8.4% of the electricity is used between 7AM to 9AM. Since the metered data comes from a social housing building, socioeconomic factors might also explain why the DHW use has no morning peak, but a more balanced consumption during the day with occupants adapting different schedules. However, since this study used data from a single building, it is not currently possible to assess whether this discrepancy is really caused by the social housing aspect of the building or by other factors. The shape of the measured electricity consumption profile is similar to the one simulated for the weekend (the models predicts that to 7AM to 9AM period is responsible for 6.7% of electricity use during the weekend).

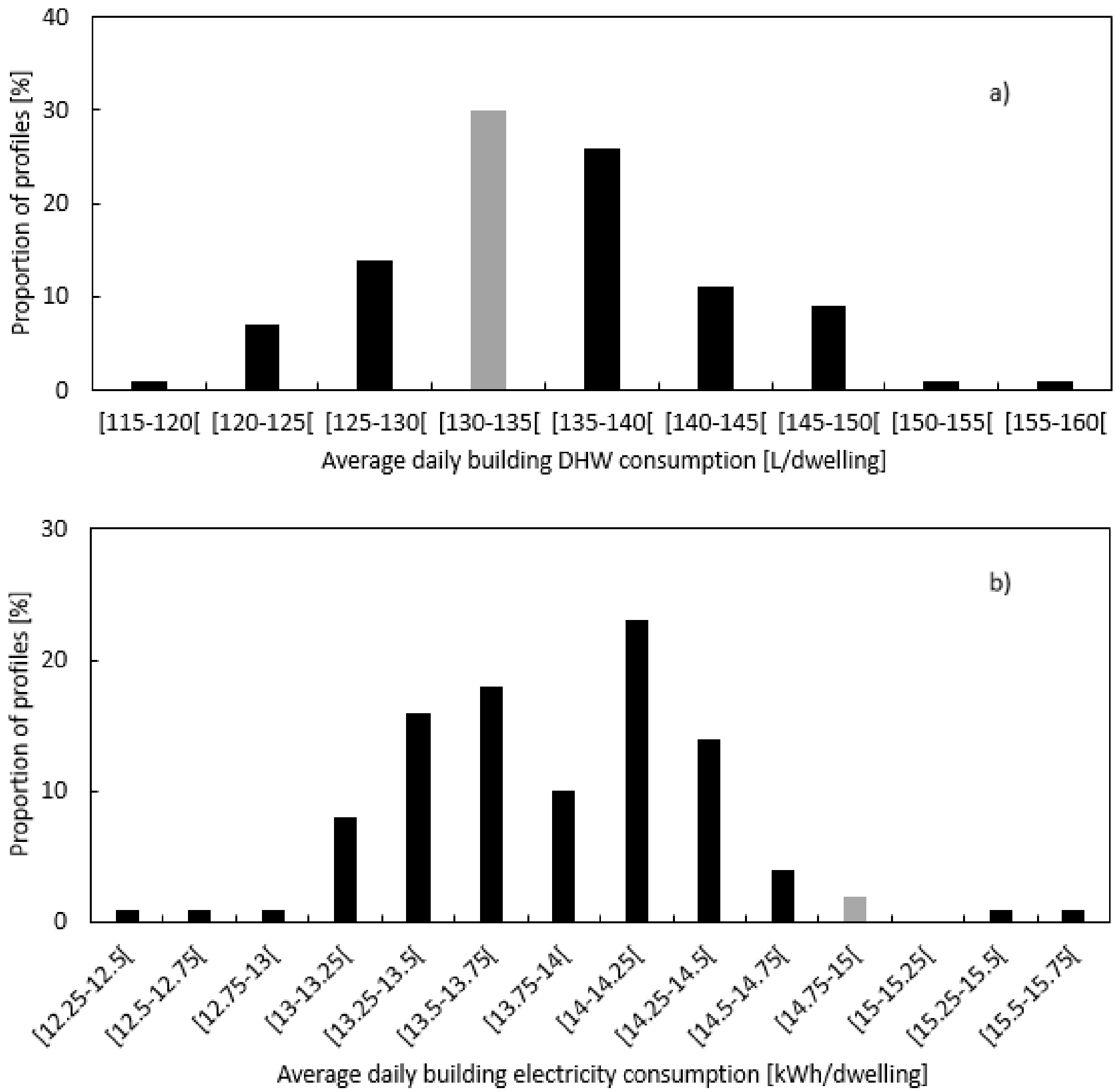

Figure 6: Distribution of the a) average DHW and b) electricity daily consumption per dwelling obtained after 100 simulations. Shaded bar represents the cluster in which the monitored building falls into.

Notwithstanding this difference in the morning, peak heights are roughly the same in the simulated and measured datasets. Regression coefficients between the measured and generated time series are $R^2 = 0.855$ for DHW and $R^2 = 0.890$ for electricity consumption. Moreover, the differences seen between the measured and simulated DHW use profiles do not lead to errors for the sizing of the hot water system [24]. It can thus be concluded that the aggregated daily behavior of the model fits reasonably well with the measurements. If the goal was to represent more closely the case study building, one would need to scale down the probability of DHW and electricity demand events in the morning.

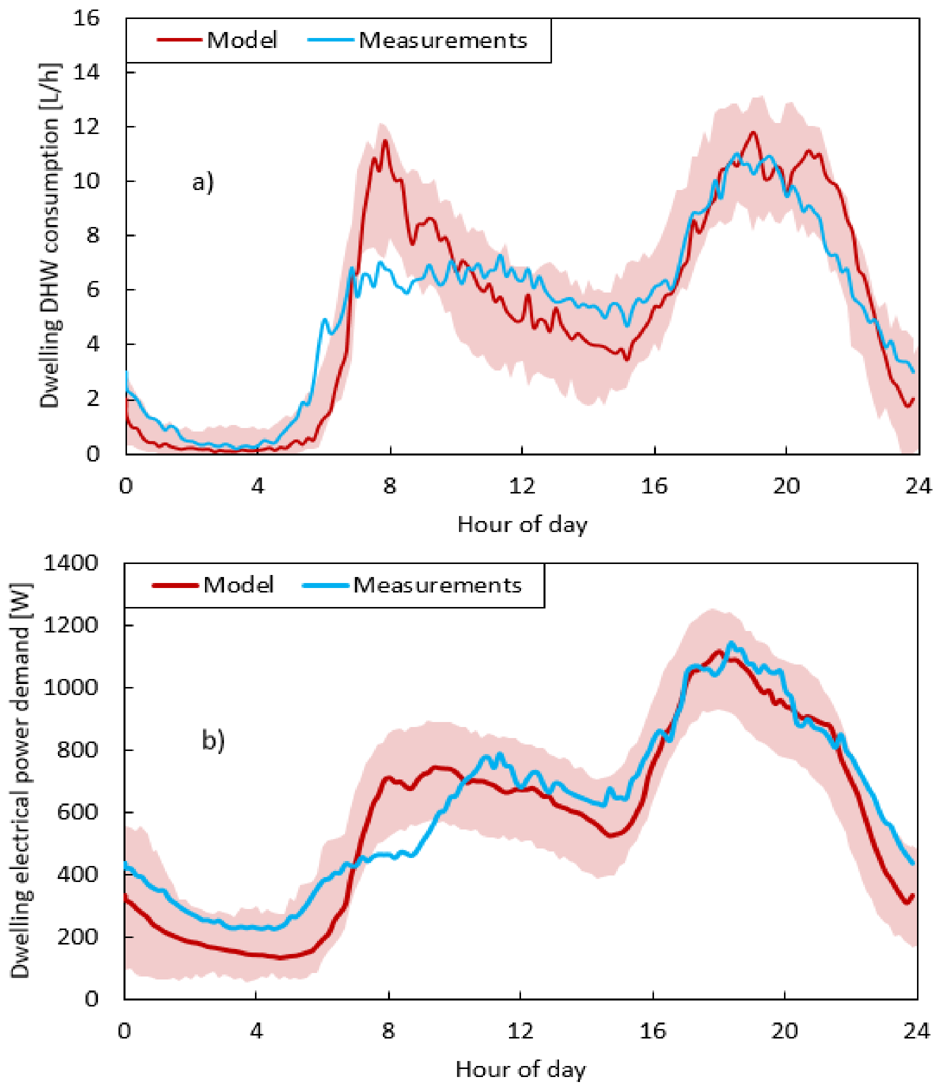

Figure 7: Average daily (weekdays and weekend days combined) a) DHW and b) electricity use by simulated and measured dwellings over a year from 100 simulations. Shaded areas represent the range prescribed by the 5th and 95th percentiles obtained from the 100 simulated profiles.

3.2 Disaggregated demand

The variability in consumption between different dwellings generated by the model is examined in contrast with the one observed in the real building. Among the 100 simulated building profiles, the one that produced the level of DHW consumption and electricity that were the closest to the real building was selected and is analyzed here. The measured standard deviation of daily consumption between the 40 dwellings is 95.2 litres for hot water and 5.93 kWh for electricity. In the selected simulated profiles, these values respectively are 42.5 litres and 6.60 kWh, meaning that although the variability for electricity consumption is accurate, the model is conservative in terms of variability among households for domestic hot water. Further work to obtain more data about this variability would be helpful to get an improved representation. The goodness-of-fit between the observed distribution and the one predicted by the model was assessed with Mann-Whitney test. The computed p-values are $3.52 \cdot 10^{-5}$ for the hot water distribution and 0.357 for electricity use. At a significance level of 95%, these values mean that the model fits with observed data for electricity consumption, but not for DHW. This is confirmed by Fig. 8 which displays separately the consumption of every measured and simulated dwelling. In the case of DHW (Fig. 8a), contrarily to the simulation results, there are several very-heavy users in the building as well as low-consumption occupants.

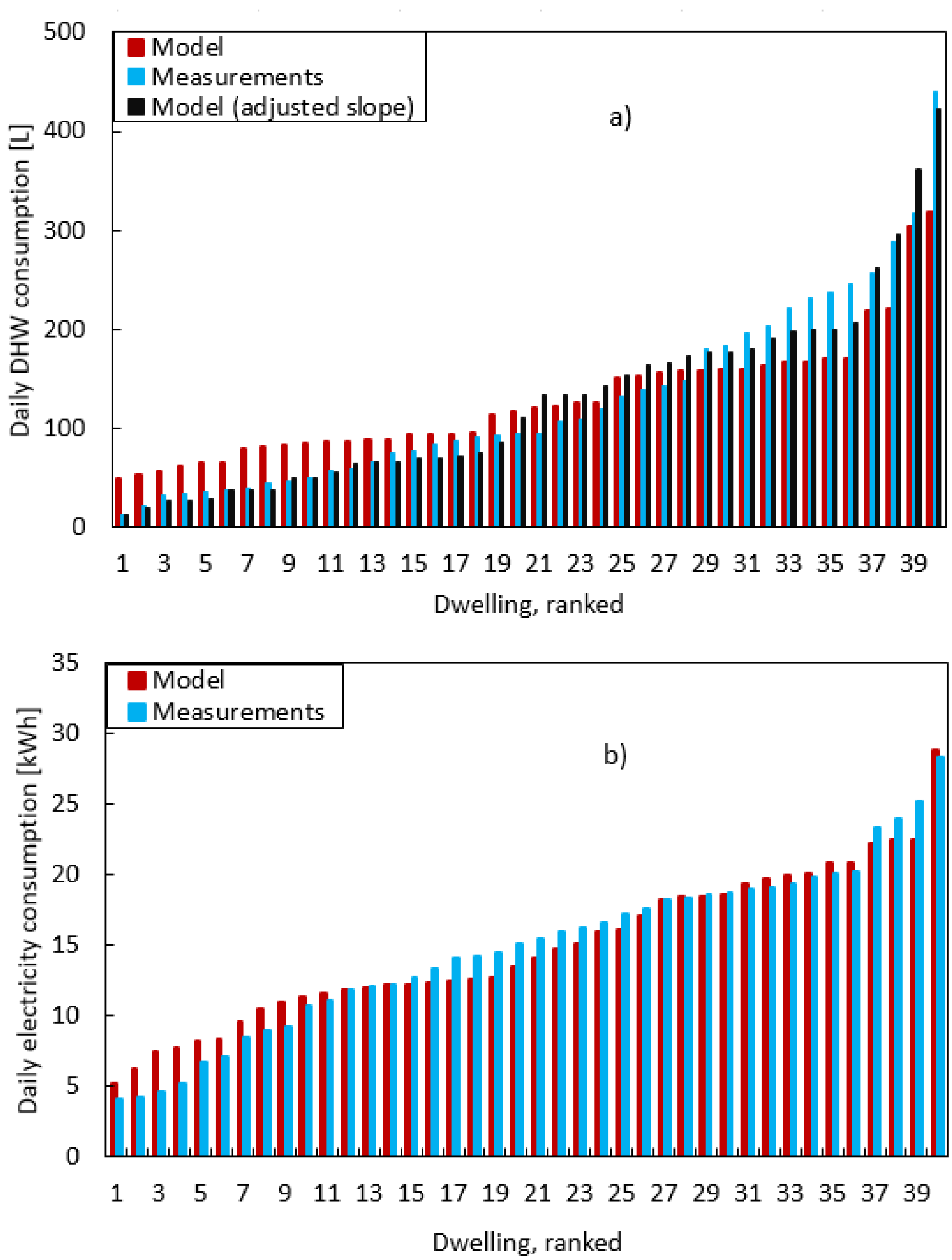

Figure 8: Average daily a) DHW and b) electricity profiles from 100 simulations compared to the one measured from the case study building.

To identify the reason behind this disparity, the DHW consumption of dwellings was plotted in Fig. 9 by separating them according to their household sizes. Fig. 9 also offers best fit lines computed from linear regression for the estimation of DHW demand with the household size. The diversity of consumption around the linear regressions is slightly underestimated by the model. The larger diversity in the measured data appears to be mostly caused by the larger impact of household size on hot water use. A comparison of the linear regression equation reveals that the household size has twice as big an influence in the monitored data (slope of 55 litres per person) than in the simulations (27 litres per person). Consequently, there is an important difference in consumption between dwellings with low and high household sizes, explaining the larger variability. The test was re-run with a slope of 55 litres per person prescribed in the model. This modification significantly increased the goodness-of-fit between the distribution seen in the monitored building and the one predicted by the model. The new p-value of 0.331, indicating that both distributions fit at a significance level of 95%. Black bars in Fig 8a represent the interhousehold distribution obtained with the new slope – it can be seen that it follows the measured distribution more closely than the simulated distribution generated with the previous slope. A slope of 55 litres per occupant is larger than those found elsewhere. Studies have reported a slope of 26 L/person in the UK [43] and of 35 [45] and 39 L/person [42] in surveys made in Canada. The presence of numerous families with young children might once again be responsible for this difference. Larger households are those with young children, who consume more hot water, hence the increase of the slope. The slope used in the model can easily be readjusted by users.

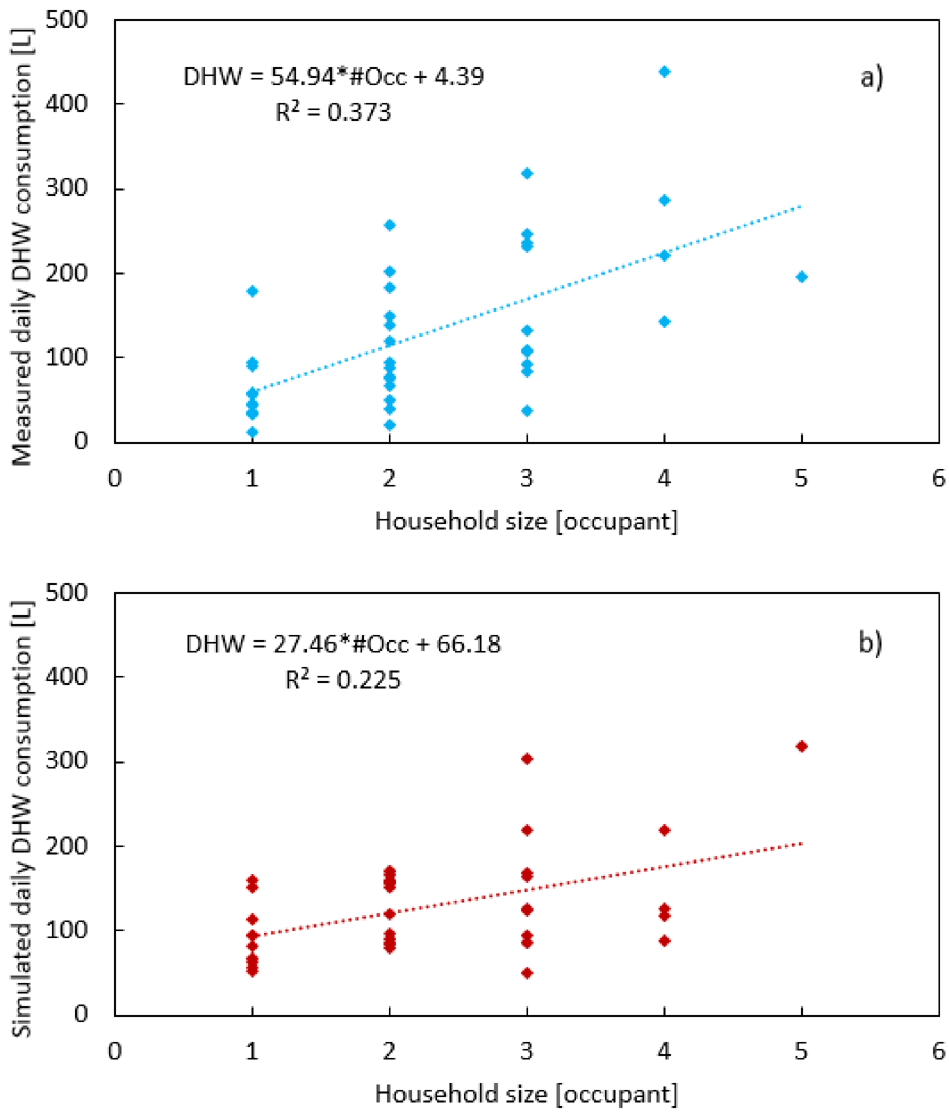

Figure 9: Consumption of DHW as a function of household size according to a) measurements and b) simulations.

Fig. 10a offers a visual depiction of how all simulated DHW consumption profiles compared with measured data. The first column on the left that is separated from the others is the measured profile, from the lowest-consuming dwelling to the highest. The other columns represent the 100 profiles generated from simulation, after the change of the DHW-per-occupant slope, and ranked by total DHW consumption. Note that for the sake of visibility, the colorbar is topped at 300 L per day. Fig. 10b presents the inverse cumulative distribution function of daily DHW demand from metered data (blue curve) and simulations (shaded areas). The black shaded area is the variations seen from the 5th

and 95th percentiles observed from the 100 simulated profiles before the change of the slope and the red one is obtained after the change, showing that the change of slope was beneficial. When expressed on a per capita basis, simulated daily DHW consumption vary from ~31 L per day per person to ~114 L per day per person, from low-use to high-use consumers. This result is coherent with literature, e.g. ASHRAE handbook [61].

Figs. 10c and 10d are respectively the electricity consumption equivalent of Figs. 10a and 10b. Again, a maximum value of 30 kWh is used in Fig. 10c to improve visibility of the variations. Fig. 10d reveals that the 100 simulated profiles all match fairly well with the measured building profile, except for a slight divergence for the low-consuming households (those set in the lowest 10%).

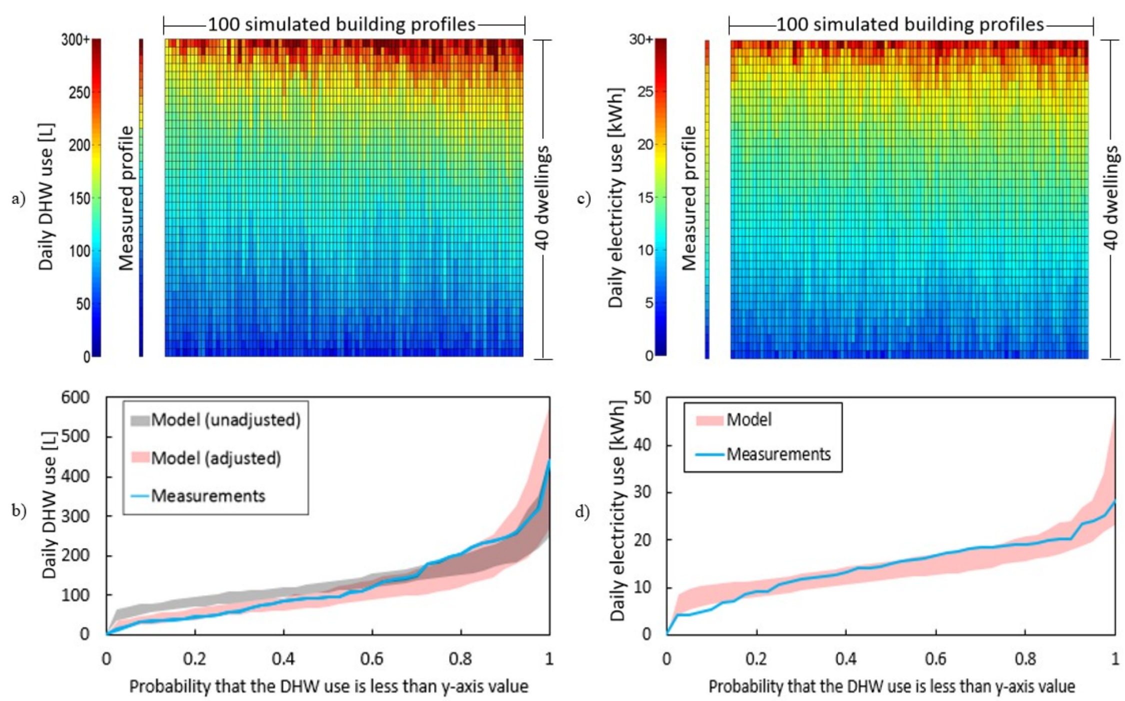

Figure 10: a) Average dwelling daily DHW consumption for all measured and simulated profiles (x-axis: the 100 profiles, y-axis: the 40 dwellings). b) Inverse cumulative probability function of the DHW consumption of a dwelling from measurements and simulations. c) Average dwelling daily electricity consumption for all measured and simulated profiles (x-axis: the 100 profiles, y-axis: the 40 dwellings). d) Inverse cumulative probability function of the DHW consumption of a dwelling from measurements and simulations.

Mann-Whitney goodness-of-fit tests yields acceptable fit at a significance level of 95% for 97 of the 100 slope-adjusted DHW profiles (from 3 out of 100 with an unadjusted slope) and 92 of the 100 electricity profiles.

Patterns of residential energy consumption exhibit some stochastic variation in multiple dimensions. In addition to modeling diversity in consumption among buildings, day-to-day variations must also be modelled for each dwelling. People do not consume the same quantity of energy day after day. Figs. 11 and 12 exhibits the day-to-day variability of the measured and simulated dwellings. Centerlines in the boxes represent the median day of consumption, edges of the boxes the first and third quartiles and the whiskers show the position of the 5th and 95th percentiles. Note that for electricity, Fig. 12 could only be generated for the eight dwellings whose electricity consumption is measured as daily consumption for the other apartments is unavailable. For both DHW and electricity, the model generated day-to-day variability that is nearly constant for all dwellings as shown by the similar length of the boxes and whiskers in Figs. 11 and 12. A different pattern is seen for the measured data, in which day-to-day variability is fluctuating from a dwelling to another. Some households consume a very consistent volume of DHW day after day and others do not. For example, in the case of electricity demand, dwellings #3 and #4 have a nearly identical median day, but the narrower box evidences that the consumption in dwelling #3 is much more consistent than in dwelling #4.

The average day-to-day standard deviation for DHW is 65.9 litres in the validation data and 57.9 litres in the simulation profile; while for electricity, these values are 6.13 and 4.48 kWh respectively. Therefore, it appears that the model generates less day-to-day variation for electricity and hot water use than occurs in reality. No factor was introduced in the model to force diversity of consumption between different days for a single dwelling. This diversity is driven by the probabilistic nature of the occupant behavior model. It appears that this is not sufficient and that another factor would be valuable to enhance the day-to-day variability of a simulated dwelling. Such factor could be drawn from a PDF and could vary every day.

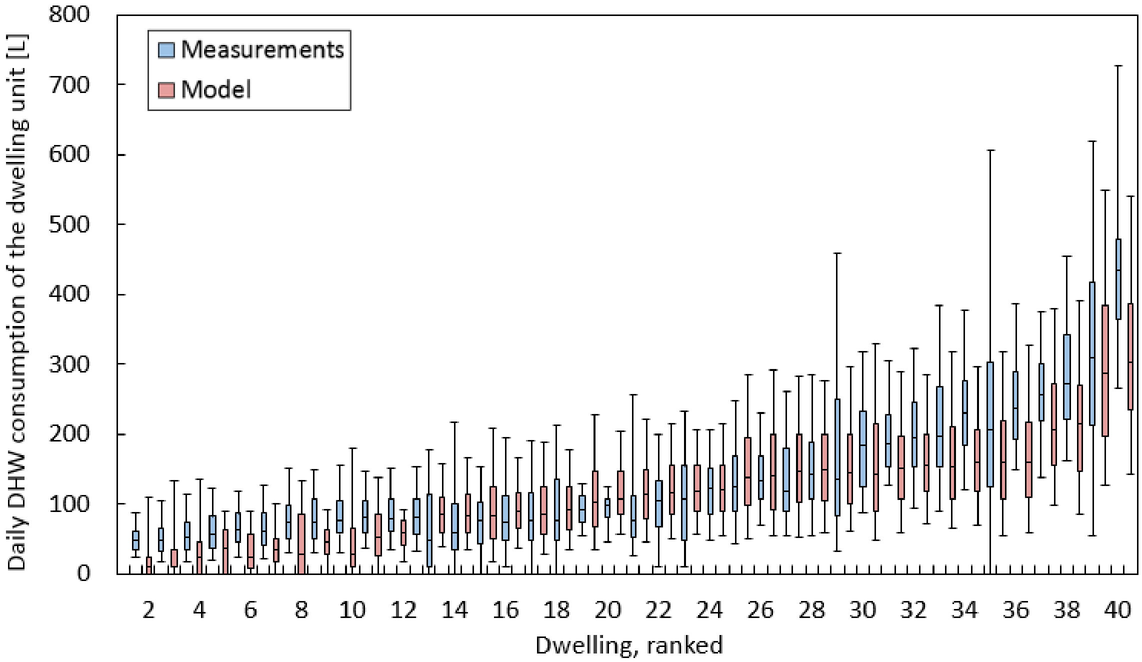

Figure 11: Measured and simulated day-to-day variability of DHW consumption.

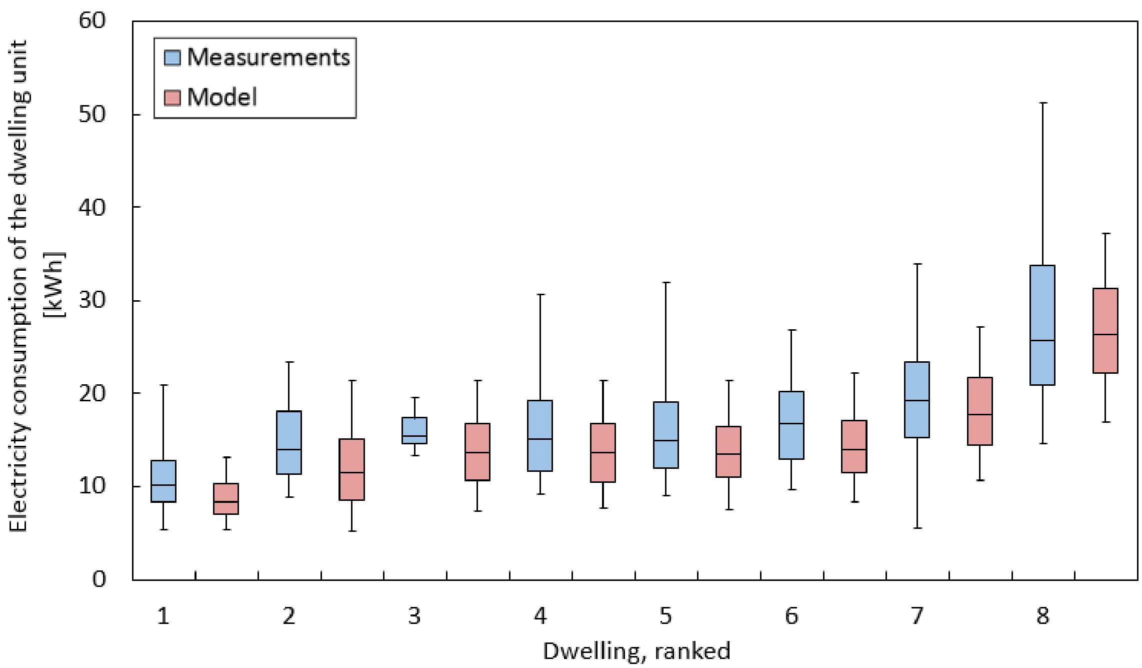

Figure 12: Measured and simulated day-to-day variability of electricity consumption.

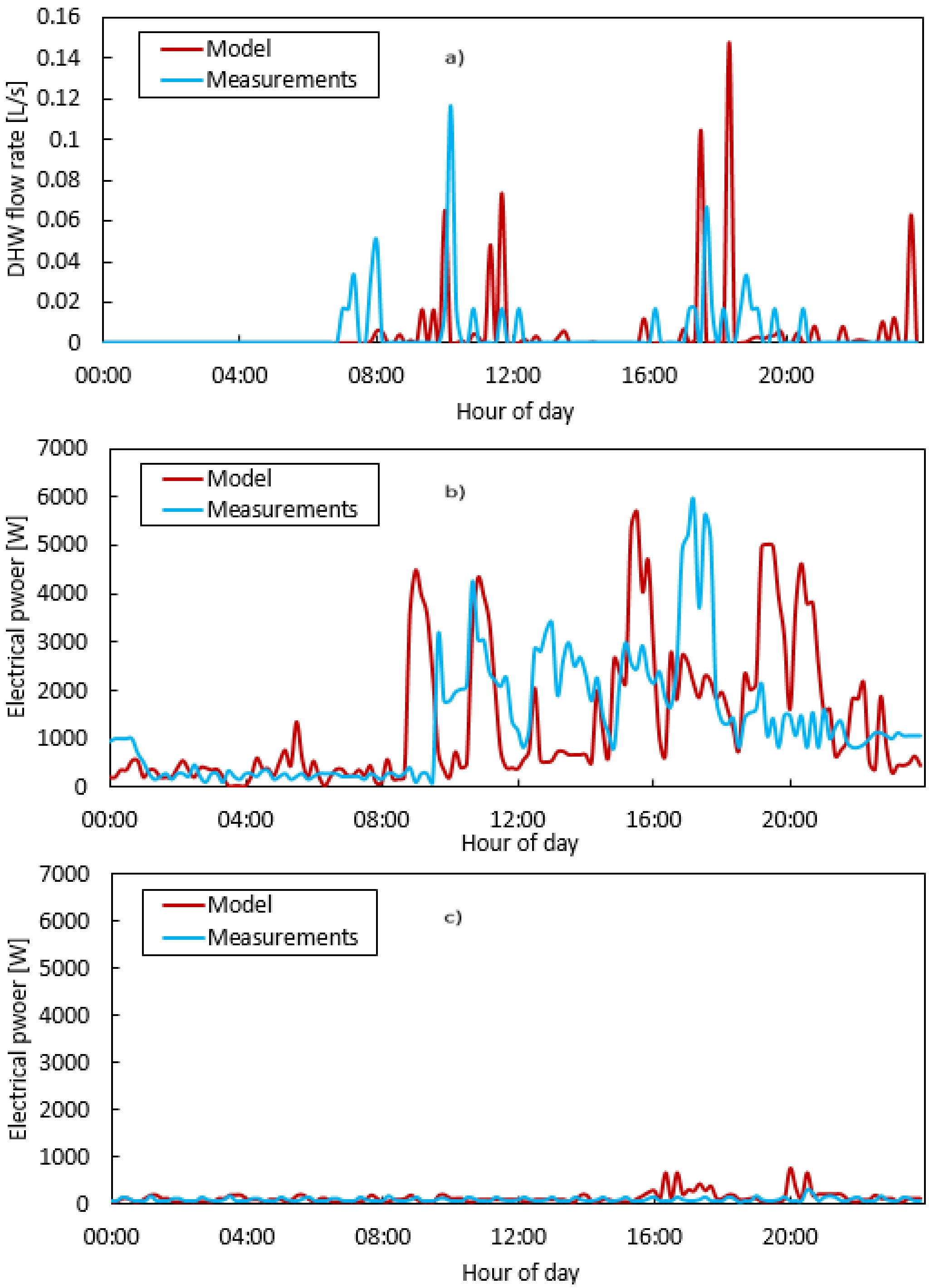

Figure 13: Simulated and measured daily schedule of a) DWH use during the highest day of consumption b) electricity use during the highest day of consumption and c) electricity use during the lowest day of consumption for a selected dwelling. Minimal DHW use during the lowest day of consumption is not shown since it yielded zero consumption for both simulations and measurements.

Figure 13 illustrates the consumption schedules during individual days for one selected dwelling. The dwelling was randomly selected from the simulation profiles and then it was paired with a dwelling from the monitored building that yielded a similar level of consumption. Fig. 13a presents the maximum day of DHW consumption (a total volume of 370.0 litres was consumed during that day in measurements, 399.3 in simulations), Fig. 13b the maximum day for electricity consumption (32.3 kWh in measurements, 32.4 in simulations) and Fig. 13c the minimum day for electricity consumption (2.2 kWh in measurements, 3.0 in simulations). The day that had the lowest use of DHW is not displayed since in both the model and validation datasets this day had zero consumption of hot water. The purpose of Fig. 13 is merely to show the profile trends – a perfect match between the curves is not expected. The DHW curves have a similar behavior: zero consumption for most of the days along with ten to twenty spontaneous short consumption events. Peaks of consumption related to an occurring event have comparable magnitude. The peak heights are also similar for electricity consumption. Curves for this part of the model show that electricity use oscillates when the dwelling is in "standby mode". When occupants are truly using electrical appliances, the power demand increases greatly. A zoom on Fig. 13c exposes that the standby power is smaller in the model (41 W) than it was in the monitored dwelling (60 W). This gives a reason for the underestimation of consumption during the night in the aggregated profile (see Fig. 7) since an underestimation of 19 W throughout the day translates into an energy consumption of 0.46 kWh/day per dwelling. Looking back at Fig. 6, considering such an offset would move the measured electricity use closer to the average calculated from the simulations. This offset could be explained by the choice of electrical appliances in the dwellings. Nevertheless, extreme days yield similar total amount of energy use between the simulated and the measured apartment. The overall trends were adequately reproduced, demonstrating the capacity of the model to generate realistic daily profiles.

Overall, there is a good fit in terms of aggregated and disaggregated patterns between the profiles that are generated by the model and the measurements made in a real building. Yet, there remains discrepancies that suggest that more data has to be collected for further improving the model. For example, a 'day-to-day variability' factor which control the consumption level of every day could be useful for the model, but no study on the day-to-day variability in consumption can be found in literature and thus it is not possible to obtain an appropriate PDF from which this factor could be drawn. Additionally, one could question the relevance of adding such a factor as it would slow down the computations without necessary adding information that is important for building design. Another way of improving the model could be the characterisation of different user types via a differentiation of behavior. The model could assign to each dwelling the type of DHW users (morning versus evening users) that live in it and then adjust hot water events PDFs accordingly. To do so, one needs to know the proportion of people that consume more water in the morning, which is very difficult to quantify.

3.3 Effects of changes on accuracy of model predictions

To create a unified probabilistic model for the simulation of occupant behavior in residential buildings, several changes were applied to already existing models as described before. This section verifies how each of these changes influences the accuracy of the simulations. Three indicators were chosen to assess the performance of the occupant behavior model. First, the relative difference of overall consumption between the case study building and the average obtained from 100 simulations of the building was computed:

$$I_{cons} = \frac{\frac{1}{n}\sum_{i=1}^{n} Q_i - Q_m}{Q_m} \times 100\% \tag{7}$$

where Q_m is the average daily measured quantity, Q_i is the average daily simulated quantity for the ith generated profile and n is the number of simulated building profiles (n = 100 here). The second performance indicator is related to the timings of consumption and looks at the average daily schedule of consumption:

$$I_{sched} = 100\% \times \frac{Q_m}{\bar{q}_m} \sqrt{\frac{\sum_{j=1}^{144}\left(\frac{q_{j,m}}{Q_m} - \frac{1}{n}\sum_{i=1}^{n}\frac{q_{ij}}{Q_i}\right)}{144-1}} \tag{8}$$

where $q_{j,m}$ is the average measured rate of consumption for the jth time step of the day and q_{ij} the average simulated rate of consumption obtained from the ith generated profile. The 144 value in Eq. (8) comes from the fact that there are 144 time steps during a day when using a 10-min frequency. The average rate of consumption are divided by the average daily consumption in order to ensure that changes in overall consumption (which are already measured by the first indicator) do not also influence the second performance index. The final indicator is the discrepancy between the measured and simulated coefficient of dwelling-to-dwelling variation:

$$I_{dwellings} = \frac{\frac{1}{n}\sum_{i=1}^{n} CV_i - CV_m}{CV_m} \times 100\% \tag{9}$$

The coefficient of dwelling-to-dwelling variation is defined as the standard deviation of the overall consumption of dwellings in a building divided by the average consumption of the building. Once again, dividing the standard deviation by the average consumption ensures that discrepancy in overall consumption will not be reflected in this indicator. The three performance indices were computed after each change was cumulatively applied to the occupant behavior model for both DHW and electricity consumption. The computed indices are presented in Table 2. The blue cases in Table 2 were implemented before this validation test to represent where changes are expected to have an effect on the model, e.g. the first change (scaling for apartment or detached houses) is only expected to influence the overall consumption of electricity predicted by the model.

All three indicators are error functions, so low values for the indicators indicate better performance. The figures in Table 2 demonstrate that the changes applied were greatly beneficial for the prediction of DHW and electricity use in terms of overall consumption in the building and of dwelling-to-dwelling variability. For the DHW section, adjusting the daily hot water use from 27 to 55 L per occupant as done during the validation reduced the underestimation of dwelling-to-dwelling variability from 37.2 to 9.4%. Although an underestimation of 37.2% as initially obtained after applying the "type of consumer" parameter appears unsatisfactory, the introduced parameter still significantly reduced the error on the dwelling-to-dwelling variability as it was set at an underestimation of 83.9% in the original model. The introduced modifications did not have a high impact on the timings of the hot water consumption, merely reducing I_{sched} from 30.4 to 24.2% for DHW and from 18.6 to 15.1% for electricity. This is explained by the fact that the changes brought to the occupancy part of the model had no significant impacts on the simulation, with the three performance indices staying nearly unchanged before and after the introduction of those changes. It appears that the two scale factors related to occupancy were not able to correct the fact the schedules obtained from British lifestyle was used to simulate the behavior of Canadians. The fact that a social housing building was used for the validation may also explain this lack of improvement as occupancy behavior in a dwelling might change according the socioeconomical status of its occupants. More data on active occupancy and activity schedule need to be available if one wants to improve the prediction of the scheduling of hot water and electricity events in the occupant behavior model.

Table 2. Performance of DHW and electricity prediction after applying various changes applied to already existing occupant behavior models.

#	Section of the model	Change	DHW			Electricity		
			I_{cons} [%]	I_{sched} [%]	$I_{dwellings}$ [%]	I_{cons} [%]	I_{sched} [%]	$I_{dwellings}$ [%]
0		-	72.5	30.4	-83.9	-41.4	18.6	-73.6
1	Electricity	Scale for type of dwelling	72.6	30.6	-83.7	-66.6	18.5	-74.7
2		Scale for electricity appliances (UK to Canada)	72.5	30.6	-83.9	-14.4	15.5	-74.7
3		Scale for occupant activities (UK to Canada)	72.4	30.5	-83.8	-7.6	16.7	-72.4
4		Electricity/Household size slope	72.6	30.7	-83.8	-8.4	16.8	-26.1
5		Add the “Type of consumer” parameter	72.4	30.6	-83.8	-8.4	16.8	-7.1
6	DHW	Link DHW with occupancy	24.1	24.4	-92.1	-7.9	16.8	-6.7
7		Scale for hot water appliances (USA to Canada)	22.4	23.9	-92.3	-8.9	16.8	-7.4
8		Scale for low-flow devices	3.2	23.6	-92.7	-8.3	16.8	-6.8
9		Add the “Type of consumer” parameter	2.9	23.5	-37.2	-7.8	16.8	-7.9
10		Adjusted the slope from 27 to 55 L/(day*person)	2.6	23.4	-9.4	-8.2	16.8	-7.3
11	Occupancy	Scale for active occupancy (UK to Canada)	2.2	23.9	-9.1	-5.9	15.6	-5.2
12		Add the “Type of occupant” parameter	2.7	24.2	-9.5	-6.4	15.1	-2.0

4. Conclusions

A strategy to create a unified probabilistic occupant behavior model for Canadian multi-residential buildings was proposed and tested. This strategy merges multiple recognized models built in different parts of the world. Since occupants in different countries could have different behaviors, scaling is necessary to adapt already existing models to specific locations worldwide. This was possible since Canada, US and UK share similar occupant behavior patterns. Modifications were also necessary to make sure that the outputs from the occupant behaviors models were coherent. In this paper, this idea has been shown to be possible for Canadian lifestyle. The scaling was based on national aggregated statistics about time-use, DHW demand and electricity consumption of Canadians. These data are more accessible in most countries than the large datasets required to build a new occupant behavior model. Therefore, it appears easier to scale a model from one country to another than to create a completely new model. The behaviors considered in the developed model are occupancy, domestic hot water use and consumption of electricity. The model has a time resolution of 10 minutes. Four already existing models were merged and scaled in this new model: Richardson's active occupancy and domestic electricity use models, Hendron's DHW profile generator and Armstrong's model for the simulation of stochastic lighting loads in dwellings. It was found that additional scale factors are needed to ensure that there is a significant diversity in consumption between different dwellings and that the level of consumption is coherent with the household size of the dwellings.

The model predictions were validated with measured data from a multi-residential building in Canada. The validation section of this work shows that the aggregated simulation and measurement results agree with one another better than previous models. Even though every building has unique differences that are difficult to predict without very detailed knowledge about the residents' behavior, the remaining discrepancies were relatively small and could be explained by a lack of data (e.g. data concerning the DHW consumption of young families). Despite minor differences, the total consumption of the building falls into the range predicted by the model, and the average daily profiles have similar patterns. Most of the differences between the model and measurements might be explained by the large number of young families in the real building. The difference in consumption between the dwellings is well replicated for electricity but not for DHW, for which it underestimated. Further analyses have shown that this underestimation is mainly caused by the misrepresentation of the relation between DHW consumption and household sizes. Household size is more important for DHW demand than usual in the monitored building, again likely due to the numerous young families. As for the day-to-day diversity of consumption for an apartment, while its representation was adequate for DHW consumption, the diversity for electricity demand is too narrow when compared with validation data. An additional scale factor that infers different levels of consumption for each day could fix this shortcoming. This could be important in certain applications; for example, in evaluating the instantaneous pairing of PV systems with buildingelectricity demand. New studies on the variations of electricity consumption between different days for one household would be necessary to implement such a factor and is recommended for further work. Nevertheless, the newly developed model was shown to offer better performance than the original models for the simulation of DHW and electricity consumption in a multi-residential building in Canada.

The model was developed with the objective of being coupled with building simulation software. The model could also be used in several disciplines such as sociology, psychology, grid design, urban logistics and many others. With respect to energy assessment models, the generated profiles could directly provide occupancy, DHW and electricity use time series to the building numerical model, which is crucial for the calculations of internal gains and of the overall energy demand of the building. To estimate internal gains generated by the occupants themselves or for performing calculations of air quality and contaminants diffusion, it would be beneficial to know when they are sleeping in the building. The model currently does not discern between being away from the building and being in the building, but sleeping. Therefore, a possible improvement would of a third state (sleeping) in the occupancy model. Also, socioeconomic factors were not directly considered in the version of the model presented in this paper. In preconstruction simulations, it could be difficult to know the household composition of a dwelling. Instead of weighting the model for age, gender, salary and other social parameters, it was thus decided to use scale factors drawn from probability density functions created to simulate the variability in consumption related to those parameters. Considering socioeconomic factors (age, salary, energy price, education…) could increase the accuracy of the model, in particular when one wants to simulate a specific and existing building for which this information is available. For instance, the energy consumption in the case study building was more balanced than predictions from the unified model during the day with no peak of consumption during the morning. This discrepancy might be explained by the young population of the building and/or by its social housing aspect. However, considering these factors would require significantly more data as the observations made in this paper are derived from a single case study building. More monitoring studies on occupant behavior in different types of residential buildings are needed to further increase our understanding on this topic.

The existing base models used to create the united model presented in this paper were developed in Canada, United Kingdom and United States. Although differences in occupant behavior are observed between these countries, one could argue that their socioeconomical environment are similar, which eased the process to adapt the models for Canada. The methodology would need to be tested with countries where residents have substantially different domestic hot water use or electricity consumption patterns. For example, these differences might come from work schedules (e.g., variation of the number of hours spent at work vs at home in different countries), energy price (e.g., the energy price structure in a country might influence the way people consume energy), climate (e.g., number of hours spent inside versus outside, use of artificial vs natural lighting, etc.), and so on. The extent to which the approach used in the paper could be extended to countries with very different occupation behaviors is yet an open question. To minimize bias in the scheduling of occupancy and of energy events, using occupancy data or models from a specific country will always be preferable than using scaled data from another country, but when this option is unavailable, the scale strategy seems to provide satisfying results for the generation of realistic energy use profiles.

Acknowledgements

The authors are grateful to Natural Sciences and Engineering Research Council of Canada for the financial support through its IRC and CRD programs (IRCPJ 461745-12 and RDCPJ 445200-12) as well as the industrial partners of the NSERC industrial chair on eco-responsible wood construction (CIRCERB). The first author is also supported by the *Fonds de recherché du Québec – Nature et technologies* (FRQNT). This work was possible thanks to a Canadian Queen Elizabeth II Diamond Jubilee Scholarship which allowed the first author to spent four months at the University of Bath (UK). The authors are also grateful to the *Société d'Habitation du Québec* (SHQ) for providing the data for the residential building presented in Section 4. Ramallo-González would like to thank the program Saavedra Fajardo (20035/SF/16) funded by *Consejería de Educación y Universidades of CARM*, via *Fundación Séneca-Agencia de Cinecia y Tecnología de la Región de Murcia*.

Declarations of interest: none

Appendix A

The appendix includes Tables A1-A3.

References

[1] IEA International Energy Agency, "Energy technology perspectives 2016 - Towards sustainable urban energy systems," Paris, 2016.

[2] A. F. Emery and C. J. Kippenhan, "A long term study of residential home heating consumption and the effect of occupant behavior on homes in the Pacific Northwest constructed according to improved thermal standards," *Energy*, vol. 31, no. 5, pp. 677–693, Apr. 2006.

[3] B. Boardman, "Home truths: a low-carbon strategy to reduce UK housing emissions by 80% by 2050," University of Oxfords' Environmental Change Institute, Nov. 2007.

[4] P. de Wilde and D. Coley, "The implications of a changing climate for buildings," *Build. Environ.*, vol. 55, pp. 1–7, Sep. 2012.

[5] V. Fabi, R. V. Andersen, S. Corgnati, and B. W. Olesen, "Occupants' window opening behaviour: A literature review of factors influencing occupant behaviour and models," *Build. Environ.*, vol. 58, pp. 188–198, Dec. 2012.

[6] S. A. Sadeghi, P. Karava, I. Konstantzos, and A. Tzempelikos, "Occupant interactions with shading and lighting systems using different control interfaces: A pilot field study," *Build. Environ.*, vol. 97, pp. 177–195, Feb. 2016.

[7] M. Schweiker and M. Shukuya, "Comparison of theoretical and statistical models of air-conditioning-unit usage behaviour in a residential setting under Japanese climatic conditions," *Build. Environ.*, vol. 44, no. 10, pp. 2137–2149, Oct. 2009.

[8] D. Yan and T. Hong, “EBC Annex 66 - Definition and simulation of occupant behavior in buildings,” International Energy Agency, Birmingham, UK, ISBN 978-0-9996964-7-7, 2018.
[9] P. Hoes, J. L. M. Hensen, M. G. L. C. Loomans, B. de Vries, and D. Bourgeois, “User behavior in whole building simulation,” *Energy Build.*, vol. 41, no. 3, pp. 295–302, Mar. 2009.
[10] D. Yan *et al.*, “Occupant behavior modeling for building performance simulation: Current state and future challenges,” *Energy Build.*, vol. 107, pp. 264–278, Nov. 2015.
[11] B. Bordass, R. Cohen, and J. Field, “Energy performance of non-domestic buildings: closing the credibility gap,” presented at the Building performance congress, 2004.
[12] J. Widén and E. Wäckelgård, “A high-resolution stochastic model of domestic activity patterns and electricity demand,” *Appl. Energy*, vol. 87, no. 6, pp. 1880–1892, Jun. 2010.
[13] M. Schweiker, F. Haldi, M. Shukuya, and D. Robinson, “Verification of stochastic models of window opening behaviour for residential buildings,” *J. Build. Perform. Simul.*, vol. 5, no. 1, pp. 55–74, Jan. 2012.
[14] U. Wilke, F. Haldi, J.-L. Scartezzini, and D. Robinson, “A bottom-up stochastic model to predict building occupants’ time-dependent activities,” *Build. Environ.*, vol. 60, pp. 254–264, Feb. 2013.
[15] D. Fischer, T. Wolf, J. Scherer, and B. Wille-Haussmann, “A stochastic bottom-up model for space heating and domestic hot water load profiles for German households,” *Energy Build.*, vol. 124, pp. 120–128, Jul. 2016.
[16] A. P. Ramallo-González, T. S. Blight, and D. A. Coley, “New optimisation methodology to uncover robust low energy designs that accounts for occupant behaviour or other unknowns,” *J. Build. Eng.*, vol. 2, pp. 59–68, Jun. 2015.
[17] I. Richardson, M. Thomson, and D. Infield, “A high-resolution domestic building occupancy model for energy demand simulations,” *Energy Build.*, vol. 40, no. 8, pp. 1560–1566, 2008.
[18] R. Hendron, J. Burch, and G. Barker, “Tool for generating realistic residential hot water event schedules,” *Natl. Renew. Energy Lab. Gold. CO Pap. No NRELCP-550-47685*, 2010.
[19] I. Richardson, M. Thomson, D. Infield, and C. Clifford, “Domestic electricity use: A high-resolution energy demand model,” *Energy Build.*, vol. 42, no. 10, pp. 1878–1887, Oct. 2010.
[20] M. Schweiker, M. Hawighorst, and A. Wagner, “The influence of personality traits on occupant behavioural patterns,” *Energy Build.*, vol. 131, pp. 63–75, Nov. 2016.
[21] R. McKenna, L. Hofmann, E. Merkel, W. Fichtner, and N. Strachan, “Analysing socioeconomic diversity and scaling effects on residential electricity load profiles in the context of low carbon technology uptake,” *Energy Policy*, vol. 97, pp. 13–26, Oct. 2016.
[22] Centre for Time Use Research, “Multinational Time Use Study.” [Online]. Available: https://www.timeuse.org/mtus.
[23] Mathworks, *MATLAB*. 2014.
[24] J. Rouleau, L. Gosselin, and A. Ramallo-González, “Sizing methodology for domestic hot water systems based on simulated occupant behaviour,” presented at the 2017 ASHRAE Annual Conference, Long Beach, USA, 2017.
[25] J. Rouleau, A. P. Ramallo-González, and L. Gosselin, “Towards a comprehensive tool to model occupant behaviour for dwellings that combines domestic hot water with active occupancy,” presented at the 14th Conference of International Building Performance Simulation Association, San Francisco, USA, 2017.
[26] S. Brooks, A. Gelman, G. Jones, and X.-L. Meng, Eds., *Handbook of Markov Chain Monte Carlo*. Boca Raton: Chapman and Hall/CRC, 2011.
[27] Ipsos-RSL and Office for National Statistics, “United Kingdom Time Use Survey, 2000 (Computer file),” UK Data Archive (distributor), Colchester, Essex, Sep. 2003.

[28] Statistics Canada, “Canadian households in 2011: Type and growth,” Catalogue no. 98-312-X2011003, 2011.
[29] M. Muratori, M. C. Roberts, R. Sioshansi, V. Marano, and G. Rizzoni, “A highly resolved modeling technique to simulate residential power demand,” *Appl. Energy*, vol. 107, pp. 465–473, Jul. 2013.
[30] M. A. López-Rodríguez, I. Santiago, D. Trillo-Montero, J. Torriti, and A. Moreno-Munoz, “Analysis and modeling of active occupancy of the residential sector in Spain: An indicator of residential electricity consumption,” *Energy Policy*, vol. 62, pp. 742–751, Nov. 2013.
[31] J. Widén, A. M. Nilsson, and E. Wäckelgård, “A combined Markov-chain and bottom-up approach to modelling of domestic lighting demand,” *Energy Build.*, vol. 41, no. 10, pp. 1001–1012, Oct. 2009.
[32] Statistics Canada, “General Social Survey on Time Use - Overview of the Time Use of Canadians 2005,” Ottawa, Catalogue no. 12F0090XIE, Jul. 2006.
[33] E. McKenna, M. Krawczynski, and M. Thomson, “Four-state domestic building occupancy model for energy demand simulations,” *Energy Build.*, vol. 96, pp. 30–39, Jun. 2015.
[34] D. Aerts, J. Minnen, I. Glorieux, I. Wouters, and F. Descamps, “A method for the identification and modelling of realistic domestic occupancy sequences for building energy demand simulations and peer comparison,” *Build. Environ.*, vol. 75, pp. 67–78, May 2014.
[35] L. Gelažanskas and K. A. A. Gamage, “Forecasting hot water consumption in dwellings using artificial neural networks,” in *2015 IEEE 5th International Conference on Power Engineering, Energy and Electrical Drives (POWERENG)*, 2015, pp. 410–415.
[36] A. Lomet, F. Suard, and D. Chèze, “Statistical Modeling for Real Domestic Hot Water Consumption Forecasting,” *Energy Procedia*, vol. 70, pp. 379–387, May 2015.
[37] C. Sandels, J. Widén, and L. Nordström, “Forecasting household consumer electricity load profiles with a combined physical and behavioral approach,” *Appl. Energy*, vol. 131, pp. 267–278, Oct. 2014.
[38] J. Widén, M. Lundh, I. Vassileva, E. Dahlquist, K. Ellegård, and E. Wäckelgård, “Constructing load profiles for household electricity and hot water from time-use data—Modelling approach and validation,” *Energy Build.*, vol. 41, no. 7, pp. 753–768, Jul. 2009.
[39] D. Popescu and E. Serban, “Simulation of domestic hot-water consumption using time-series models,” presented at the 6th IASME/WSEAS International Conference on Heat Transer, Thermal Engineering and Environment, Rhodes, Greece, 2008.
[40] V. Bakker, A. Molderink, J. L. Hurink, and G. J. M. Smit, “Domestic Heat Demand Prediction Using Neural Networks,” in *2008 19th International Conference on Systems Engineering*, 2008, pp. 189–194.
[41] R. Hendron and C. Engrebrecht, “Building America Research Benchmark Definition,” NREL, Golden, CO, TP-550-47246, Dec. 2009.
[42] D. George, N. S. Pearre, and L. G. Swan, “High resolution measured domestic hot water consumption of Canadian homes,” *Energy Build.*, vol. 109, pp. 304–315, Dec. 2015.
[43] Energy saving trust, “Measurement of Domestic Hot Water Consumption in Dwellings,” 2008.
[44] D. S. Parker and P. Fairey, “Estimating Daily Domestic Hot-Water Use in North American Homes,” *ASHRAE Trans.*, vol. 121, p. 258, 2015.
[45] Natural Resources Canada, *HOT2000*. Ottawa, 2016.
[46] K. Adye, N. S. Pearre, and L. G. Swan, “On-site Assessment Summary Report: House Characteristics and Water and Energy Savings,” Halifax Regional Municipality, Halifax, 2014.

[47] M. Chitnis and L. C. Hunt, “Modelling UK household expenditure: economic versus noneconomic drivers,” *Appl. Econ. Lett.*, vol. 18, no. 8, pp. 753–767, Jun. 2011.
[48] J. L. Harris and L.-M. Liu, “Dynamic structural analysis and forecasting of residential electricity consumption,” *Int. J. Forecast.*, vol. 9, no. 4, pp. 437–455, Dec. 1993.
[49] D. Fischer, A. Härtl, and B. Wille-Haussmann, “Model for electric load profiles with high time resolution for German households,” *Energy Build.*, vol. 92, pp. 170–179, Apr. 2015.
[50] A. Capasso, W. Grattieri, R. Lamedica, and A. Prudenzi, “A bottom-up approach to residential load modeling,” *IEEE Trans. Power Syst.*, vol. 9, no. 2, pp. 957–964, May 1994.
[51] R. Baetens *et al.*, “Assessing electrical bottlenecks at feeder level for residential net zero-energy buildings by integrated system simulation,” *Appl. Energy*, vol. 96, pp. 74–83, Aug. 2012.
[52] A. Marszal-Pomianowska, P. Heiselberg, and O. Kalyanova Larsen, “Household electricity demand profiles – A high-resolution load model to facilitate modelling of energy flexible buildings,” *Energy*, vol. 103, pp. 487–501, May 2016.
[53] M. M. Armstrong, M. C. Swinton, H. Ribberink, I. Beausoleil-Morrison, and J. Millette, “Synthetically derived profiles for representing occupant-driven electric loads in Canadian housing,” *J. Build. Perform. Simul.*, vol. 2, no. 1, pp. 15–30, Mar. 2009.
[54] R. Subbiah, “An activity-based energy demand modeling framework for buildings: A bottom-up approach,” Thesis, Virginia Tech, 2013.
[55] Statistics Canada, “Households and the environment: Energy use,” Ottawa, 1920–7522, Apr. 2010.
[56] National Resources Canada, “Survey of house energy use - Detailed statistical report,” Ottawa, 978-1-100-23468–7, 2011.
[57] Natural Resources Canada, “Water Heater Guide.” Oct-2012.
[58] D. George, “New profiles of occupancy driven appliance, lighting, plug loads and hot water use for residenial sector energy demand modeling,” Dalhousie University, Halifax, Nova Scotia, 2016.
[59] D. Godoy-Shimizu, J. Palmer, and N. Terry, “What Can We Learn from the Household Electricity Survey?,” *Buildings*, vol. 4, no. 4, pp. 737–761, Oct. 2014.
[60] F. M. Andersen, M. Baldini, L. G. Hansen, and C. L. Jensen, “Households’ hourly electricity consumption and peak demand in Denmark,” *Appl. Energy*, vol. 208, pp. 607–619, Dec. 2017.
[61] ASHRAE, “Service Water Heating,” in *ASHRAE Handbook - HVAC Applications*, Atlanta, GA, 2015.

Table A1. Daily amount of time spent on various household activities for the average person [26], [31].

Activities	Canadian data [min]	British data [min]
Active occupancy	492	527
Cooking	42	37
Watching TV	126	85
Household work	73	57

Table A2. Aggregated daily DHW use per dwelling for five water appliances [17], [54], [60].

Hot water appliances	Canadian data [L/day]	American data [L/day]
Shower	59	73
Bath	40	18
Sink	81	65
Clothes washer	36	24
Dishwasher	9	15
Total consumption	225	195

Table A3. Specifications used by the model for each appliance to compute their operating schedule and energy consumption [18], [50].

Appliance	Activity	Operating Power [W]	Standby power [W]	Event length [min]	Probability of use	Annual consumption in Canada [kWh/year]	Annual consumption in the UK [kWh/year]
Refrigerator	None	265	0	20	0.1902	801	87
Freezer	None	263	0	20	0.1916	614	277
Desktop computer	Occupant	250	5	300	0.0023	749	247
Laptop computer	Occupant	130	0	300	0.0016	156	-
Stereo	Occupant	120	9	60	0.07858	153	80
Coffee maker	Occupant	900	0	3	0.1330	130	-
Kettle	Occupant	1500	1	3	0.1662	225	157
Lighting [141 m^2]	Occupant	-	0	-	-	2030	715
Dishwasher	DHW	467	0	35	-	94	91
Clothes washer	DHW	505	1	30	-	99	149
TV 1	Watching TV	100	3	73	0.0631	99	236
TV 2	Watching TV	100	3	73	0.0635	99	140
TV receiver box	Watching TV	40	2	73	0.1104	63	128
Exhaust fan	Cooking	250	0	30	0.2035	90	-
Hot plate	Cooking	1250	1	16	0.1715	219	128
Microwave	Cooking	1500	2	30	0.0658	197	66
Toaster	Cooking	1200	0	3	0.2598	58	-
Range	Cooking	1600	3	43	0.1950	770	145
Dryer	Laundry	4115	1	45	0.8892	1284	80

Hair dryer	Wash/Dress	1000	0	5	0.2042	60	-
Iron	Iron	1000	0	30	0.4675	72	16
Vacuum cleaner	House cleaning	800	0	20	0.1964	96	69